

Nemzeti könyvtárak és az OSZK MARC21 állományainak összehasonlító elemzése néhány adatmező alapján

Tanulságok

Ungváry Rudolf, Király Péter

Király Péter QR catalogue elemzőjével összevetettük a jelentősebb nemzeti könyvtárak MARC21 állományainak feldolgozását az OSZK állományával. Ebben a munkában a rekordfej (06, 07 és 17), továbbá a kódolt fizikai jellemzők (007) tartalomra vonatkozó adatait elemeztük. Ilyen nagy tömegű összehasonlító elemzésekre megfelelő informatikai eszközök hiányában eddig nem volt lehetőség. Az összehasonlító elemzéseket buborékdiagramok formájában vizualizáltuk a gyors áttekinthetőség érdekében. A szóban forgó tartalmi adatmezők felhasználását a kereséshez a mai könyvtári rendszerek mostohán kezelik. Célunk, hogy ebben a helyzetben segítséget nyújtsunk a fejlesztésben.

könyvtár, adatmező, MARC21, OSZK, katalogizálás

Bevezető

Korábbi tanulmányokban összehasonlítottuk néhány nemzeti és tudományos könyvtár katalógusaiban a MARC21 állományainak feldolgozását. Ehhez a munkához letöltöttük a vizsgált könyvtárak katalógusainak állományát, elemzésükhöz pedig a QA catalogue nevű, a MARC-rekordok minőségellenőrzését végző alkalmazást használtuk [2–4]. Az alkalmazással a MARC21-állományt a számítógépes elemzés eddig nem alkalmazott, új módszerével tártuk fel a további, tartalmi vizsgálat céljaira. Így olyan adatok tömege vált hozzáférhetővé, melyek alapján elemezhető a katalógusok MARC21 mezőinek kitöltése, e kitöltés munkájának tartalmi jellege. Mindez jelentős segítség lehet a katalógusok minőséget meghatározó strukturális jellemzők megértéséhez és a hibák feltáráshoz.

2020-ban ennek a feltárásnak a tartalomra vonatkozó adatait (043, 045, 052, 072, 080, 082, 084, 085, 505, 520, 6XX) és a velük összefüggő adatait (007, 008, 034, 041) elemeztük [9–12].

2022-ben a MARC21-katalógusok egy, a szakmai közvélemény előtt lényegében homályban maradó részével foglalkoztunk. Ennek a résznek a jellemzőit csak az egyes könyvtárak speciális feldolgo-

zói ismerik igazán, de nekik sincs rálátásuk más nagy könyvtárak hasonló adatainak kitöltési elveire. A rekordfejről (06, 07 és 17), továbbá a vezérlő mezőkről (kódolt fizikai jellemzők, 00X) van szó. Az utóbbiak közül is csak a kódolt fizikai jellemzők (007) és a meghatározott jellemzők és információs adatok (008) tartalomra vonatkozóan fontosabbnak tekintett adatait vizsgáltuk akként, hogy összehasonlítottuk a kiválasztott pozíciókon szereplő értékeket. Az eredményt a gyors áttekinthetőség érdekében vizuálisan buborékdiagramokkal szemléltettük.

E tanulmányban az összehasonlítás fókuszában az Országos Széchényi Könyvtár katalógusállománya áll. Azt vizsgáljuk, miben hasonló és miben különbözik a magyar nemzeti könyvtár katalogizációs és feldolgozási gyakorlata néhány nagyobb nemzeti könyvtárétól? A tanulságok segíthetnek a feldolgozás minőségének javításában, a katalógusok minőséget meghatározó strukturális jellemzői jobban megérthetővé, rejtett problémák láthatóvá válnak. Az állományok nagysága miatt az összehasonlítás e tanulmány keretében ugyancsak néhány kiválasztott adatmezőre terjedhetett ki. Ahogy a korábbi két tanulmányban, úgy most

is a források tartalmára vonatkozó adatelemekre helyeztük a hangsúlyt. A felhasználókat ugyanis mindig a keresett tárgykör érdekli. A tartalom szerinti keresésük eredménye alapján értékelik egy katalógus minőségét.

Az 1. táblázatban MARC21 mezőket¹ és pozíciókat vetettünk össze.

A kiválasztott rekordfej-pozíciók és vezérlőmezők adatelemei mesterséges nyelvű, enumeratív osztályozási rendszerek egyetlen karakteres értékből álló jelzetei (kódszavai, a szakmai szlengben kódjai). Velük osztályozva a dokumentumok a legkülönbébb feldolgozási mód és dokumentumtípus szerint osztályozhatók. Például: a rekordfej 07. pozícióján azt osztályozták, hogy a bibliográfiai egységet analitikusan, gyűjteményesen, kötetként, monografikusan stb. dolgozták-e föl. A rekordfej 17. pozícióján azt, hogy autopsziával, rövidítetten, előzetesen stb. A rekordfej 06. pozícióján a dokumentum fajtáját osztályozták a külső fizikai megjelenés szempontjából: nyomtatott, kéziratos, audiovizuális stb. A 007 adatmező 00 pozícióján a dokumentum fajtáját a vizuális megjelenés szempontjából osztályozták aszerint, hogy szöveg, térkép, grafika stb. (Lásd példaként a 2. és 3. diagram pozícióhoz kötött osztályait.)

A 648 és 65X adatmezők kiválasztott almezői a klasszikus, természetes nyelven alapuló információkereső nyelvek tárgyszavait képviselik. Velük analitikus részletességgel tárható föl a dokumentumok tartalma.

Megvizsgáltuk, milyen összhang áll fenn e mezők között. Például vannak-e nyilvánvaló, vagy azoknak látszó ellentmondások; teszem azt egy mű nyomtatott nyelvi dokumentum (rekordfej/06=a), ezzel szemben a kódolt fizikai jellemzők szerint nem kiveithető grafikához tartozó kollázs (007/01=c)? Vagy van-e olyan kézirat, melyet kötetként/részegységként osztályoztak? Ugyancsak kiderül az elemzésből, hogy vannak-e hibás pozícióértékek. Ezeket a diagramokban az „[érvénytelen]” kategóriába soroltuk. Ilyen nagy tömegű összehasonlító elemzésekre megfelelő informatikai eszközök hiányában eddig nem volt lehetőség.

A rekordfej és a vezérlő mezők adatai hasznos és egyszerű kereshetőségeket biztosíthatnának olyan esetekben, amikor a keresésbe a dokumentumok egyes típusait is bevonják. Ezeket az adat-

1. táblázat Az összehasonlításba bevont MARC21 mezők és pozíciók

mező	pozíció/almező	elnevezés
rekordfej	06	a dokumentum jellege (=a rekord típusa)
	07	bibliográfiai szint
	17	a leírás jellege
007	00	kódolt fizikai jellemzők, dokumentumkategória
	01	a dokumentum speciális megjelölése: szöveg
	01	a dokumentum speciális megjelölése: térkép
008	18–21	könyvek, illusztráltság
	25–27	könyvek, tartalmi jellemzők
	33	könyvek, műfaj
600		személynév tárgyszó
648	a	kronologikus tárgyszó
650	a	szaktárgyszó
651	a	földrajzi tárgyszó
653		szabadon választott tárgyszó
655	a	formai tárgyszó (dokumentumtípus)

mezőket a mai könyvtári keresőrendszerek mostohán kezelik. Gondoljuk csak meg, hogy mennyire nem egyszerű manapság egy olyan keresést megvalósítani, melynek tárgya mindazon elbeszélések, melyek illusztráltak?

A tételjes és a kódolt fizikai jellemzők osztályait a jelentőségükhöz képest a végfelhasználók csak nagyon nehézkesen tudják használni. Ezért eleve lemondanak a használatukról. Munkánk célja, hogy néhány elérhető nemzeti könyvtár vonatkozó állománya esetén a felszínre kerüljenek eme adatmezők kitöltésének jellegzetességei, és az esetleg belőlük fakadó problémák, anomáliák. Végső célunk, hogy mindezzel empirikus adatok álljanak rendelkezésre mind a katalógusok, mind a kezelőrendszerek, mind pedig nem utolsó sorban a MARC21 fejlesztéshez.

Már itt utalunk rá, hogy a rekordfej pozícióinak kitöltése kötelező, a vezérlő mezőké nem. Ennek messze ható következményei vannak. Ezeket a tanulmány végén összegeztük.

A vizsgálatba bevont nemzeti könyvtárak – a buborékdiagramokban megjelenő sorrendben – az alábbiak:

LC	Kongresszusi Könyvtár ²
DNB	Német Nemzeti Könyvtár ³
ÖNB	Oszták Nemzeti Könyvtár ⁴
NFI	Finn Nemzeti Könyvtár, ⁵ Fennica adatbázis (a nemzeti bibliográfia)
LIBRIS	Svéd Nemzeti Könyvtár és közös katalógus ⁶
BNPL	Lengyel Nemzeti Könyvtár ⁷
KBR	Belga Nemzeti Könyvtár ⁸
BL	Brit Nemzeti Könyvtár ⁹
NIZ	Izraeli Nemzeti Könyvtár ¹⁰
KB	Holland Nemzeti Könyvtár ¹¹
CSH	Cseh Nemzeti Könyvtár ¹²
OSZK	Országos Széchényi Könyvtár ¹³

Egyben minden könyvtárnak megköszönjük, hogy adataikat fölhasználhatjuk, különösen azt, ha ez nyilvánosan is lehetséges.

A buborékdiagramokban a méret nem az adott metszésponton keletkező abszolút számok, hanem az egyes könyvtárak belső rekordaránya szerint szerepel. A rekordarány tehát nem a legnagyobb könyvtár, hanem a saját állománytól függ. Abszolút számok esetében a kisebb nemzeti könyvtárak buborékjai annyira kis méretűek, hogy nehéz vizuálisan értékelhető arányokat felismerni.

A tanulmányban mezőnevek dolgában a HUN-MARC 2002. március magyar KSZ/4.1 kiadás terminológiáját követjük [1]. Ha a MARC21-ben [5] eltérés van, vagy az almező új, akkor az angol kifejezés fordítását adjuk meg [15].¹⁴

1. Rekordtípusok és bibliográfiai szint<sp>

Jól látható, hogy az összehasonlításban a bibliográfiai szint egyes típusai elsősorban a nyelvi anyagra jellemzők. Figyelemre méltó a kétdimenziós nem kivetíthető és a zenei hangzó anyagok esete. Ezek között van összefoglaló, alárendelt (kötet, részegység) és sok monografikus szintű feldolgozás. A kétdimenziós anyagok (többnyire fényképek, grafikák) esetében az OSZK-hoz hasonlóan számos más könyvtárban (LC, ÖNB, BNPL KBR, BL, NED, CSH) túlnyomórészt fényképekről van szó. A zenei hangzó anyagok pedig többnyire nem komolyzenei dalok, és a DNB-re meg a BL-re jellemzők, de kisebb arányban az OSZK-ra is. Mindez szűrőpró-

bákon alapuló becslés. Mit tartalmaznak a jelentősebb nemzeti könyvtárak belső feldolgozási szabályzatai ezeknek a dokumentumoknak az esetében? Például milyen alapértelmezéseket? Nem utolsó sorban azt se volna érdektelen tudni, hogy hogyan definiálják (ha egyáltalán) nyomtatott és hangzó zenemű, kartográfiai anyag, két dimenziós nem kivetíthető grafikák stb. esetén, hogy valami monografikus, gyűjteményes, alárendelt stb. feldolgozásúak?

Az OSZK-ban az egyes táruk számára külön kitöltési sablonok készültek, melyekben megadták, hogy alapértelmezésként hogyan kell kitölteni egy adott mezőt, almezőt, pozíciót. Amikor néhány évtizede áttértek az AMICUS könyvtári rendszerre, a sablonok előírt értékeit a rendszerben rögzítették („beégették”). Azóta kéziratos sablonokat már nem találni. Nem biztos, hogy ez szerencsés fejlemény, ha új rendszerre kell áttérni. A kitöltési szabályok ugyanis nem örökkévalóak. Ha át kell őket dolgozni, fáradságos munkával kell majd kibányászni a rendszerből a kitöltési alapértelmezéseket.

A szabályzatok szakmai nyilvánosságának hiányában az átfogó, konszenzusos, nemzetközileg egységes megoldás nemzetközi szinten nagyon korlátozott.

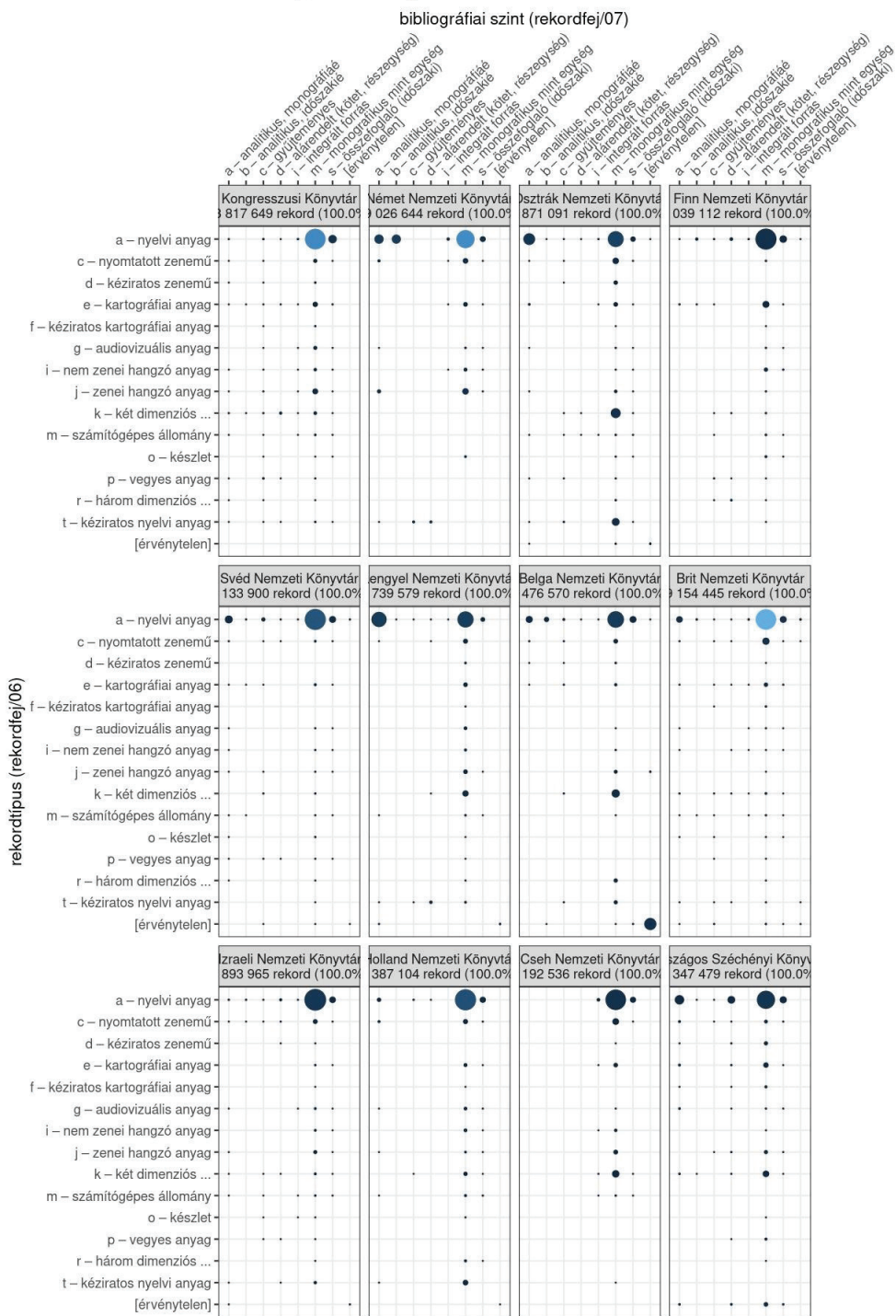
Előfordulnak más figyelemre méltó esetek. Például az LC-ben összefoglaló (időszaki) (rekordfej/07=s) a besorolása egy fényképgyűjteménynek (rekordfej/06 k) csak azért, mert 1985 óta gyűjtik benne online hozzáférhetően a képeket.¹⁵

Attól lett tehát a gyűjtemény időszaki, hogy az egyes fényképek különböző időkben készültek, ill. kerültek az adatbázisba? De akkor egy olyan cikkgyűjtemény-adatbázis, melybe különböző időkben megjelent cikkek különböző időkben folyamatosan kerülnek be, szintén időszaki volna? Kérdés, hogy az időszaki kiadványok jelenlegi meghatározása mennyire teszi ezt lehetővé? A DNB-ben például nem létezik kétdimenziós (nem kivetíthető) (azaz fénykép) mint időszaki, az ÖNB-ben igen.

1.2 Rekordtípusok és a leírás jellege (szintje)

Feltűnő, hogy általában milyen kicsi a számítógép által kezelt állomány (Rf/06=m). Az OSZK-ban (a HUNMARC-ot követve) erre az I és nem az m kódot használják, a diagramban azért nem szerepel. A maga 2000 tételes elektronikus típusú álló-

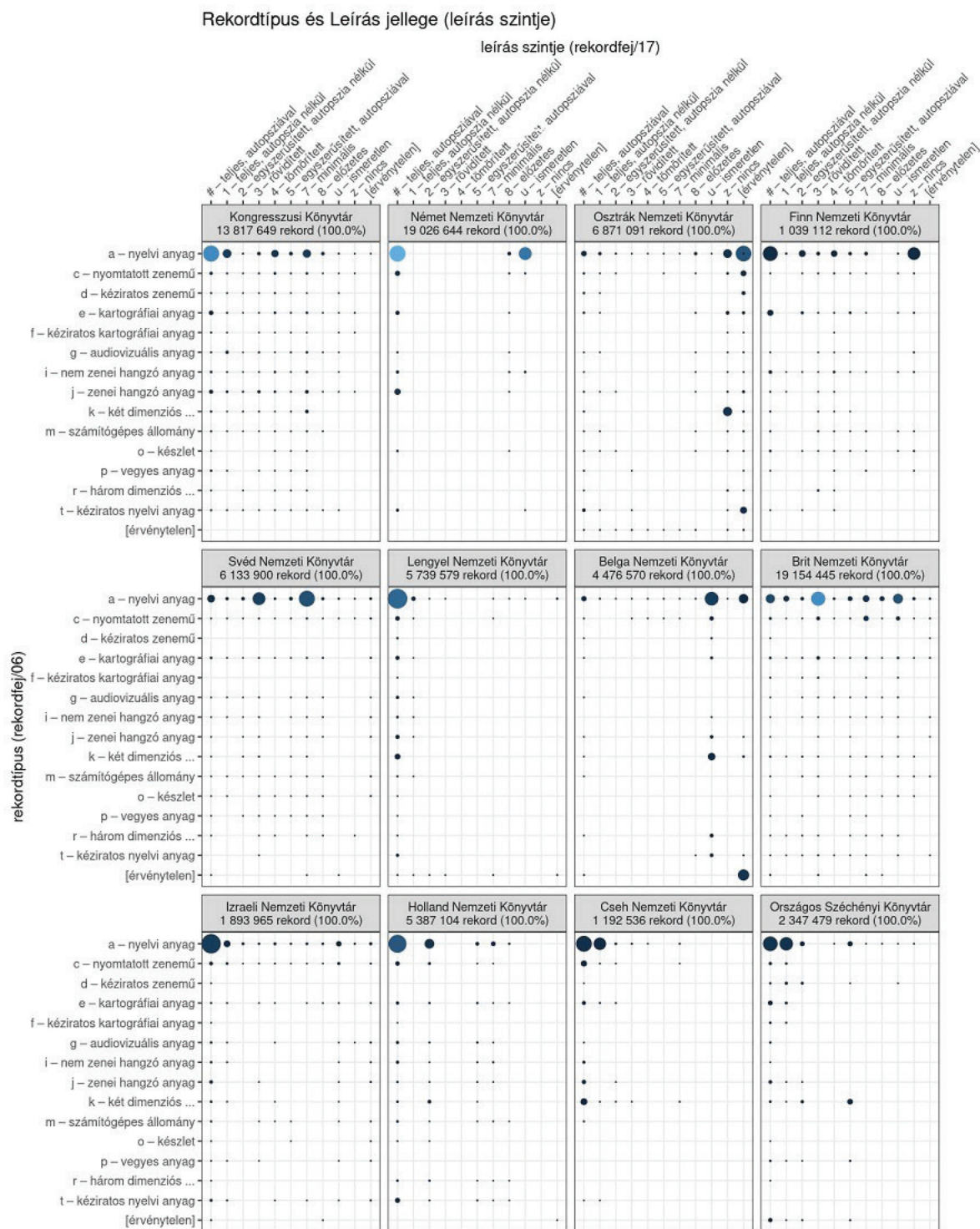
Rekordtípus és Bibliográfiai szint



1. diagram Rekordtípusok és bibliográfiai szint

mányával a hasonló nagyságú könyvtárakkal van egy sorban (ÖNB: 3.942; NBI: 1.175; összehasonlításként LC:11.017), holott az ilyen források elterjedtsége már rendkívül nagy. Erre utal, hogy az 5. diagramcsoportban a DNB-ben mennyire nagy az elektronikus minősítésű (007/00=c) dokumentumok aránya.

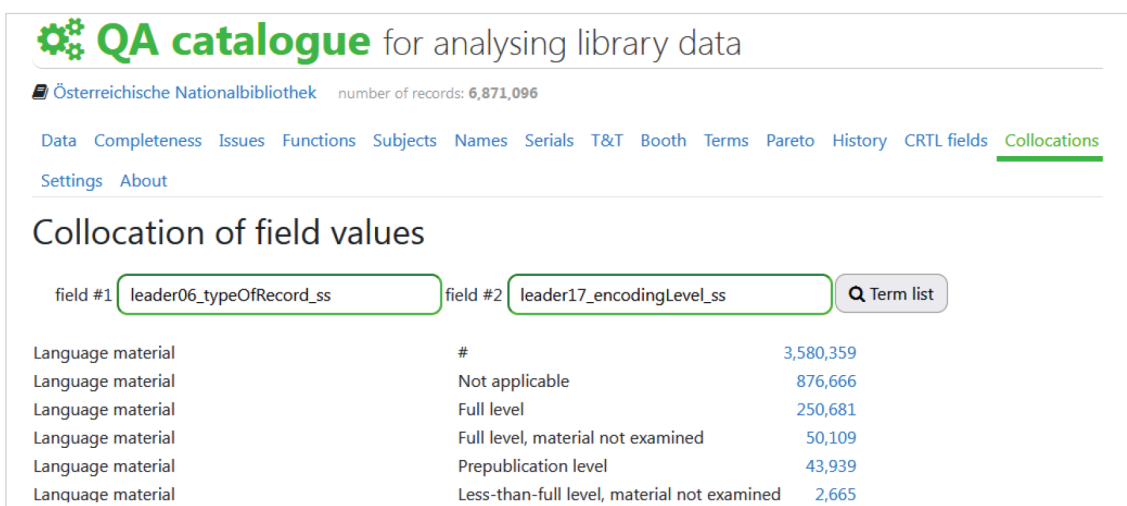
Hogy a fenti 2. diagramcsoport esetén rekordtípusként mégis annyira kevés, annak fő oka bizonyára a MARC21-formátum szabályozása. Ha ugyanis van egy jelentős szempont, amely miatt a forrás egy másik rekordtípusba is tartozik, akkor az m kód helyett az adott lényeges szempont rekordtípusá-



2. diagram Rekordtípusok és a leírás jellege (szintje)

val kell osztályozni. Így kell eljárni, ha az állomány például nyelvi anyag, grafika, kartográfiai anyag, hangzó anyag, mozgókép. Más szóval ilyen esetekben rekordtípus-adatként nem lehet több szempont szerint osztályozni, ami az egydimenziós osztályozási rendszerek alapvető hátránya.

Az LC-ben, a BL-ben és a LIBRIS-ben minden jel szerint változatosan, a különféle leírási szinteket figyelembe véve osztályoznak. A DNB-ben, ÖNB-ben és BNPL-ben feltehetően az alapértelmezések az uralkodók. Ez elmondható az OSZK-ról is. Lehet, hogy nem sok jelentőséget tulajdonítanak a rekordfej értékeinek.



1. ábra Az ÖNB # értékű rekordfej 017 pozíciói

Figyelemre méltó, hogy az ÖNB-ben a leírás szintje szempontjából a rekordtípusok nagy része a nem megengedhető kettős keresztet (#) tartalmazza (358036 rekord). Emiatt elvileg ez érvénytelen minősítésű lenne (1. ábra). Félreértésre adhatott okot, hogy a MARC-szabványban a # jelet a szóköz jelölésére használják. Ezt a tényt csak a szabvány bevezetőjében közlik, az egyes adatelemek leírásakor nem. Lehetséges, hogy a feldolgozók úgy értelmezték, hogy a kettős kereszt (#) a gyakorlati kitöltéskor is szóközt (tehát Rf/17=_) jelent? Vagy lehetséges, hogy a # jelét kitöltési szabályzatban írják elő? Netán csak hagyomány? Az ÖNB-ben mindenestre a # jelöli az üres helyet a leírás szintjén. Az érvénytelen érték a vizsgált könyvtárakban ritka, az OSZK-ban pedig egyáltalán nincs.

Figyelemre méltó az OSZK-ban a kétdimenziós grafikák (Rf/06=k) viszonylag gazdag leírás szintű feldolgozása: teljes autopsziával és anélkül (Rf/17=#, ill. 1), egyszerűsített autopsziával és anélkül (Rf/17=5, ill. 2). Ezek túlnyomórészt képes levelezőlapok.

Az ismeretlen minősítés is gyakori. Az OSZK-ban egyáltalán nincs ilyen; mint említettük, eleve azért, mert a katalogizáló rendszerük automatikusan adja meg az alapértelmezett minősítést. Mindez visszaköszön a későbbi, 3. buborékdigram-csoportban is.

A KBR-ben és a BL-ben a rekordok egy részét a leírás szintjét illetően ismeretlennek (Rf/17=u) nyilvánították. A BL-ben sok a rövidített leírás szint is (Rf/17=3).

A LIBRIS-ben a rekordok többségének leírás szintje minimális (Rf/17=u) vagy rövidített (Rf/17=3). Feltehetően léteznek mögöttük különféle szabályzatok vagy hagyományok, melyek ezt előírják.

A leírás szintje szempontjából szinte minden könyvtárban jelentős a nyelvi anyag feldolgozottsága, amint az várható volt. Az LC-ben és a DNB-ben kitűnik még a zenei hangzó anyag is. Kevésbé feltűnően, de az OSZK mellett a NIZ-re és a CSH-ra is jellemző ugyanez.

Az angol nyelvű (LC, BL) és inkább brit orientációjú könyvtárakban (az NFI és a LIBRIS, kisebb mértékben az NIZ esetén is) észrevehető inkább, hogy a többi (nem szöveges nyelvi) rekordtípusban is megjelenik a leírás szintjének minősítése. Az OSZK a leírás szintjének minősítése szempontjából a középmezőny alján foglal helyet.

A legkiegyensúlyozottabb helyzet az LC-ben van. Ott feltűnő a zenei anyag és a fényképek minimális jelenléte. Az OSZK ezzel szemben az éllavasok között van, aminek oka a gondos aprónyomtatványtári képeslap-feldolgozás és a zeneműtári feldolgozás. A zenei anyagok és a grafikák vonatkozásában hasonló a helyzet az ÖNB-ben is. Az OSZK ebből a szempontból is a középmezőny alján foglal helyet, holott maga a zenei anyag feldolgozása a Zeneműtárban alapos. A legegyszerűbb igényeket kielégítő rövid zenék (slágerek, népszerű fogyasztásra készült dalok stb.), illetve rajzos, többnyire történelmi eseményeket vagy helyeket ábrázoló grafikák minősítése a „nincs” (rekordfej/17=z). Más könyvtárakban az ilyen dokumen-

tumokat talán nem dolgozzák föl, és ezért nincs nyomuk. Holott legalábbis a fényképeknek idővel többnyire történelmi értéke lesz.

1.3 Bibliográfiai szintek leírás szintjei

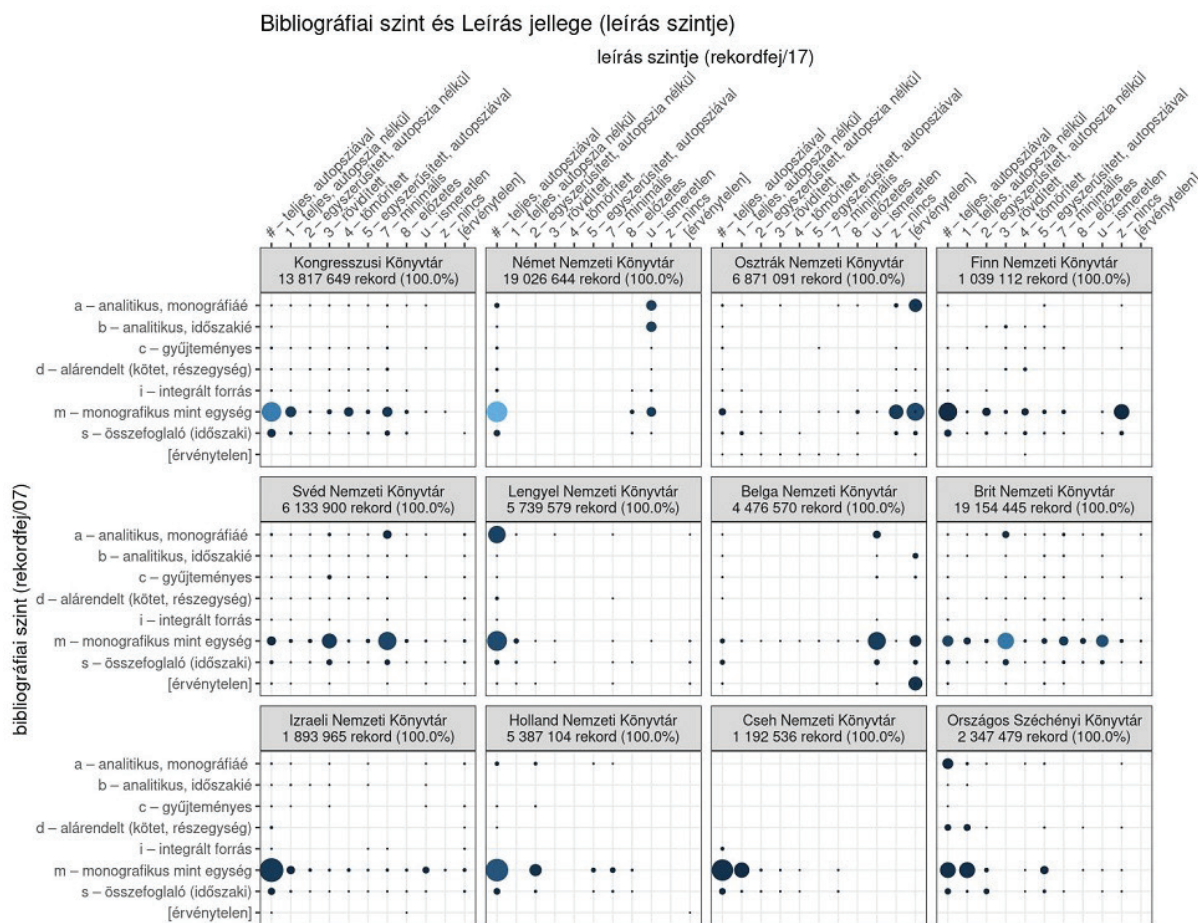
Nem meglepő, hogy a monografikus egységek leírás szintje (Rf/07=m) a legtöbb nemzeti könyvtárban következetesen a teljes leírás (Rf/17=szóköz, ill. 1). Az LC-ben és a DNB-ben a monografikus művek közel száz százalékosan teljes feldolgozásúak. Néhányban viszont, mint a LIBRIS, az ÖNB és a BL, monografikus szinten (akárcsak a többi szinten) a leírás többi szintje jelentősebb számban minimális vagy rövidített (007/17=7, ill. =3). Az OSZK-ban a rövidített és a minimális szintű leírások hiányzanak.

A kurrensen érkező könyvállományt mindig autopsziával tárják fel; ezzel szemben a többi dokumentum (jelentős részben a kartográfiai dokumentumok) autopszia nélküliek (007/17=1) vagy egyszerűsített autopsziával (007/17=5) ren-

delkeznek. E téren közel áll hozzá a CSH és a NED gyakorlata, de a két angol nyelvű nemzeti könyvtárra (LC és BR) is ez a jellemző.

Feltűnő, hogy a DNB-ben, az ÖNB-ben, a BL-ben és a KBR-ben az analitikus kiadványokat és sok monografikus egységet, mint „ismeretlent” (Rf/17=u) minősítettek. Ezekben a könyvtárakban talán inkább alapértelmezésekkel elégedtek meg. A LIBRIS-ben ehhez társul a minimális (Rf/17=7) minősítés is. Kérdés, hogy ez minek tulajdonítható? Szabályzatoknak, hagyományoknak? Az ismeretlen minősítés az OSZK-ban minimális, noha előfordul olyan rekordokban, melyeket más-honnan vesznek át.

Ahogy az előző, 2. diagramcsoportban, úgy itt is különösen feltűnő, hogy az ÖNB-ben a bibliográfiai leírás szintje ugyancsak nagy számban „érvénytelen” (Rf/17=#), és minden bizonnyal a kitöltetlenséget, az üres helyet jelöli. A KBR-ben is ez az uralkodó, ráadásul ott az érvénytelen minősítés a rekordfejben (07=[érvénytelen]) is rendkí-



3. diagram Bibliográfiai szintek leírás szintjei

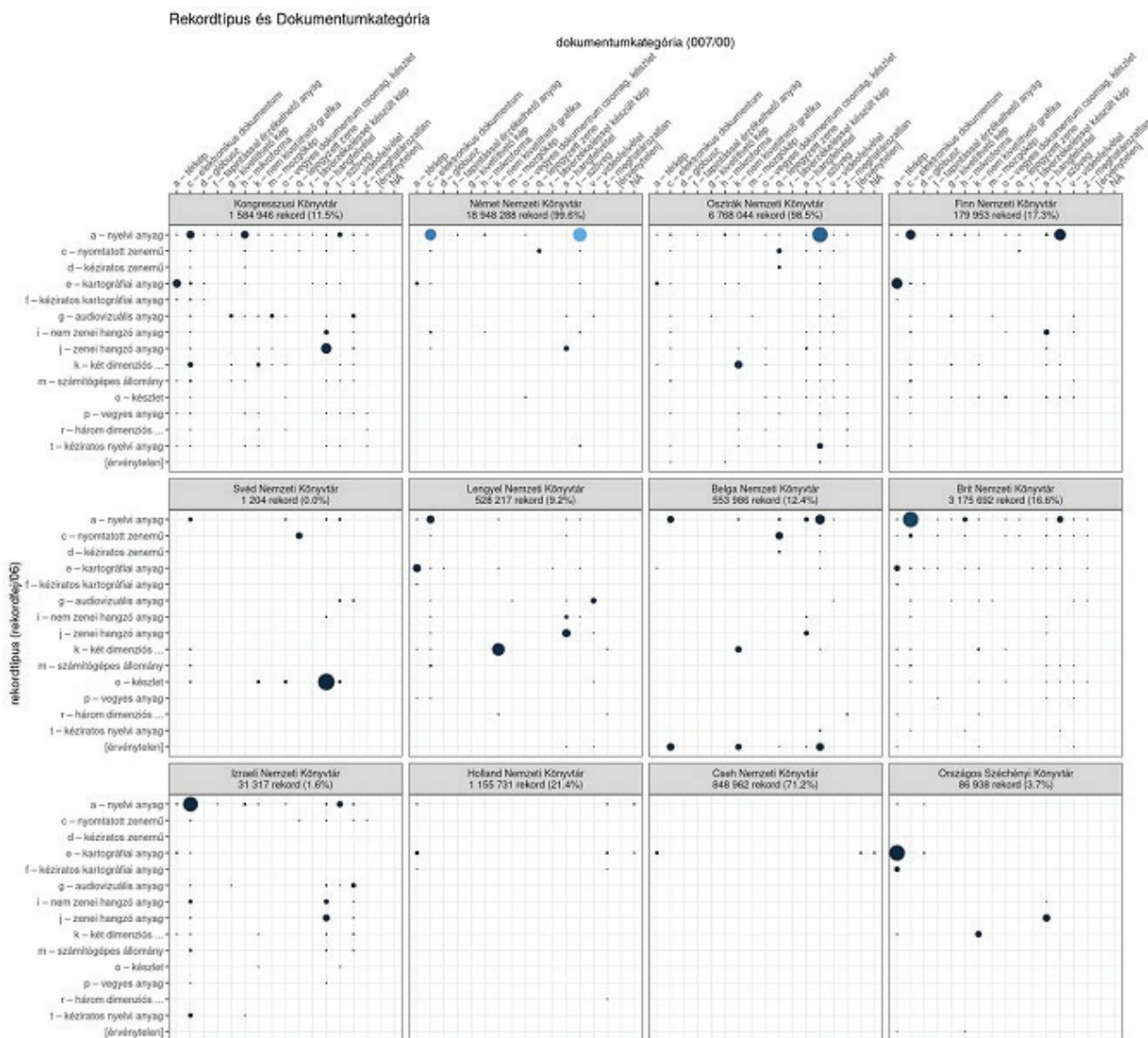
vül nagy. Mindez megint csak vagy félreértésre, vagy arra utal, hogy egy következetes, és talán belső szabályzatban dokumentált esetről van szó. Az OSZK abba a kisebbségbe tartozik, ahol semmi-féle érvénytelen nem fordul elő, ami színvonalas és jól szabályozott feldolgozásra utal.

Összefoglalóan a három diagramcsoportról elmondható, hogy *ha az összehasonlításba csak a rekordfej pozíciói vannak bevonva, akkor a feldolgozottság 100%-os*. Nem véletlenül, mivel a rekordfej minden pozíciójának kötelező értéket adni. A többi, 00X vezérlő mező esetében, ahogy ezt az alább megvizsgált 007-es és 008-as mezők mutatják, ez már koránt sincsen így, ami számos, ma még nem eléggé közismert problémát vet föl.

1.4 Rekordtípus és dokumentumkategória (Tf/6 vs. 007/00)

A diagramcsoport (ahogy a következő is) ékesen példázza, hogy mekkora eltérések vannak az egyes nemzeti könyvtárak között a nem kötelező 007-es mező kitöltésében a rekordfej 06-os pozíciójához képest.

Feltűnő, hogy az LC-ben, két német nyelvű könyvtárban (DNB, ÖNB), és több más könyvtárban (NFI, BNPL KBR, BL és NZI) viszonylag gazdag a rekordtípus Rf/06 pozíciójához viszonyítva a 007 mező dokumentumkategóriáinak kitöltöttsége. Ráadásul több könyvtárban (LC, DNB, NFI, BNPL, KBR, BL, NIZ) jelenetős az elektronikus dokumentumok nyelvi anyagként való feldolgo-



4. diagram Rekordtípus és dokumentumkategória

zása. Az OSZK-ban – ahogy néhány másik kisebb nemzeti könyvtárban is – ez teljesen hiányzik. Oka könyvek esetében nem írott szabályozás, hanem meggyökeresedett hagyomány. Csak a nyomtatott és kéziratos kartográfiai anyagok közé tartozó térképek közel teljesekek, ami nemzetközi viszonylatban is intenzív térképtári munkára utal. Jelentős még az OSZK-ban a grafikák (Rf 06=k, „kétdimenziós”) és a nem zenei hangfelvételek egymáshoz viszonyított feldolgozottsága (ahogy a BNPL-ben is). Az első az aprónyomtatványtárban a képeslapok feldolgozottságát dicséri, a hangfelvételeké pedig a Zeneműtárat. A BNPL-ben a kétdimenziós rekordtípusok (rekordfej 06=k) feldolgozottsága a mozgókép dokumentumkategóriájában (007/00=k) a leggazdagabb.

Ezek a nemzetközi összehasonlításban is megfigyelhető egyenlenségek részben következnek abból, hogy a 00X vezérlő mezők kitöltése nem kötelező.

A 007 és 008 mezők pozícióiba egykarakteres osztályozási jelzeteket kell rögzíteni. Ezek a dokumentumok típusait képviselik, tehát formai jelzetek. A pozíciók és az egykarakteres jelzetek korlátozott száma rendkívül szűkre szabott rendszer. Az osztályozónak gyakorlatilag gúzsba kötve kell tudnia táncolni. A 655-ös tárgyszó mezőben a természetes nyelvű formai tárgyszavak sokkal szélesebb választéka áll rendelkezésre.

A nem kötelező vezérlő mezők kezelése a könyvtárak belső szabályzatain és a könyvtári rendszerek (a katalogizálók által nem mindig ismert és kontrollált) beállításain múlik; esetünkben ez az OSZK-ban a sablon nevet viseli. Ezekről nem sok tudható nyilvánosan, és nagyon különbözhetnek a kitöltési előírásaik. A buborékdiagramok tanulsága szerint többnyire vagy egyáltalán nem foglalkoznak a kitöltésükkel, vagy (mint a fenti 007/00 esetben – legalábbis a két német nyelvű könyvtárban) egységesen alapértelmezéseket adnak meg. A többi esetében még ennél is kevesebb állapítható meg a külső elemző számára.

Néhány más, kisebb nemzeti könyvtárhoz hasonlóan az OSZK-ban az előbb említett, meggyökeresedett hagyomány miatt nincs minősítve a nyelvi anyag (Rf/06) a dokumentumkategóriában szövegesként (007/00=t).

Érdekes, hogy a kétdimenziós nem kivetíthető ábrázolásokat (lényegében a fényképeket, [Rf/06=k] csak az LC, az ÖNB, a BNPL), és kisebb mértékben a KBR és a BL minősíti a 007/00=k pozíción nem kivetíthető grafikaként. Pedig ez a többi könyvtárban is magától értetődő lenne. (Vagy nem, ha a kettő együttes minősítését egy belső szabályzat redundánsnak, ezért feleslegesnek minősítené, a MARC21 pedig erre vonatkozóan nem ad eligazítást.) Az OSZK-ban a szabályzat előírásait maga a katalogizáló rendszer tartalmazza.

A rekordfej 06=k pozícióján (kétdimenziós nem kivetíthető ábrázolások) és a 007/00 dokumentumkategóriák között szereplő g=kivetíthető kép és k=nem kivetíthető grafika sajátos ellentmondás is előfordul. Mind az LC-ben, mind az NFI-ben van olyan 1144 illetve 337 db rekord, melyben a kétdimenziós nem kivetíthető ábrázolások (rekordfej (06=k)) egyben a kivetíthető képek (007/00=g) besorolást is megkapták. *A feldolgozottság mértéke azonban olyan kicsi, hogy valójában nem lehet érvényes megállapításokat tenni. Ez a későbbi, 7.-12. diagramok összehasonlításai esetén már minden könyvtárra érvényes.*

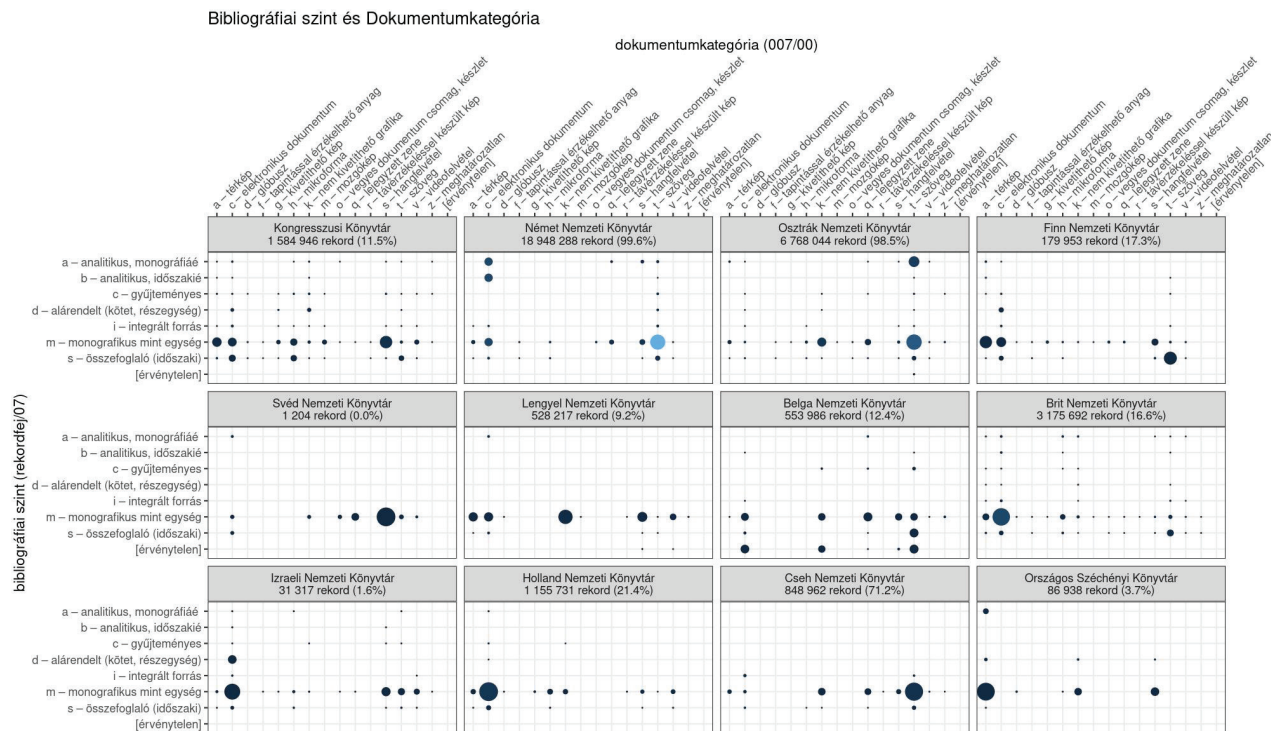
Az LC-ben a kétdimenziós nem kivetíthető ábrázolások többnyire fényképekről vagy grafikákról készült diaposzítívok. Az NFI-ben filmlapok és ugyancsak diák. Mivel több száz rekordról van szó, ez talán már nem lehet tévedés. Jobb híján osztályozhatták őket így. Inkább *azt bizonyítják, hogy a pozícióhoz kötött egykarakteres osztályozási rendszer jelzetei elégtelenek.*

1.5 Bibliográfiai szint és dokumentumkategória (Rf/07 vs. 007/00)

Hasonlóan a 3. diagramcsoporthoz, a két német nyelvű könyvtárban a feldolgozottság közel száz százalékos. A diagramok is emlékeztetnek a 3. diagramcsoporthoz, mert itt is a monografikus egységek (Rf/07=m) feldolgozása az uralkodó.

Figyelemre méltó az OSZK-ban a nem kivetíthető grafikák (007/00=k) – többnyire képeslapok – és könnyűzenei és népdal hangfelvételek (007/00=s) monografikus egységeinek feldolgozottsága.

Meglepő módon a DNB-ben, és részben a BL-ben jelentős az elektronikus dokumentumok (007/00=c) feldolgozása, főleg monografikus dokumentumok esetén, de még időszaki formában is (Rf/07=a, ill.



5. diagram Bibliográfiai szint és dokumentumkategória

b, ill. m). Ennek a többi könyvtárban csak nyomai vannak. Ahogy arra a 2. diagramcsoport esetében utaltunk, ez meglehetősen ellentmondásban van a MARC21 ama szabályozásával, hogy ha van egy jelentős szempont, amely miatt a számítógéppel kezelt forrás egy másik rekordtípusba is tartozik, akkor annak rekordtípusával kell osztályozni.

Megint csak feltételezhető, hogy a következetes gyakorlat mögött belső, dokumentált szabályozás vagy hagyomány húzódik meg.

Feltűnő, hogy a nemzeti könyvtárak közül egyedül az OSZK-ban nem minősítik a 007/00=c szerint az elektronikus dokumentumokat. Ennek oka, hogy amikor az 1980-as évek végén elkészült a rögzítést szabályozó sablon, ezek a dokumentumok még nem jelentek meg a gyűjtőkörben. Azóta a sablont nem korszerűsítették, most pedig már csak a katalogizáló rendszerben léteznek. Szerencse, hogy a 655-ös mezőben megadják az „elektronikus dokumentum” vagy annak alárendeltjeinek tárgyszavát, és ennek alapján később egyszer automatikusan érvényesíthető a 007/00 pozíció c értéke

A két német nyelvű könyvtáron kívüli könyvtárakra itt is a számszerűleg kis feldolgozottság a jellemző, ezért ezeknek a gyakorlatáról csak nagyon korlátozottan lehet következtetéseket levonni.

Mindez az összes következő diagramcsoportra érvényes.

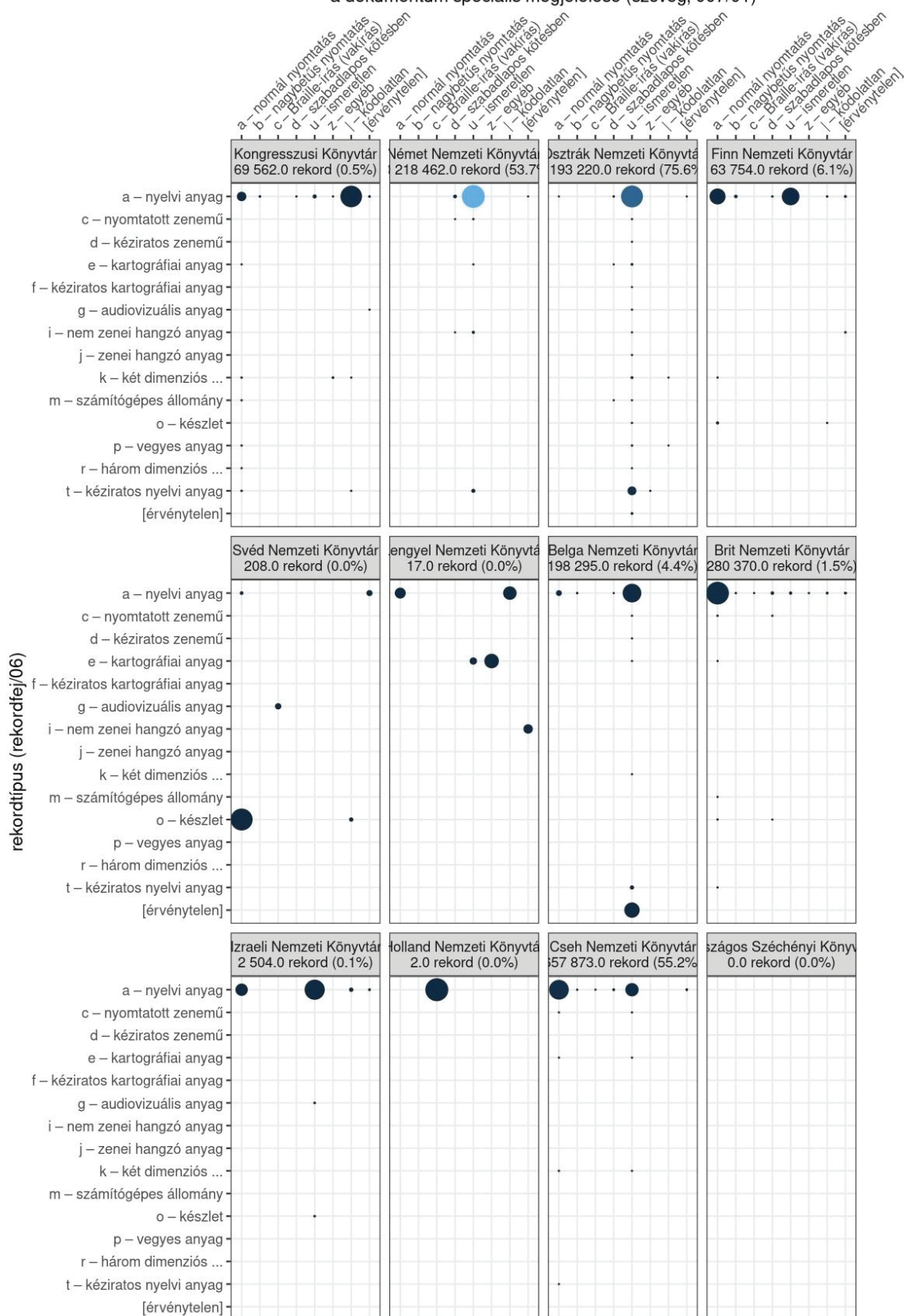
A két német nyelvű könyvtár kivételével a többinél bizonyára azért nincs a rekordok több, mint 80%-a, olykor több mint 90%-a földolgozva, mert egyszerűen kihagyják ezeket. Azaz alapértelmezésként még csak a z=meghatározatlan minősítés megadását se tartották szükségesnek megadni. Ez azért is különös, mert arra utal, mintha vagy nem is lenne belső előírás a 007 nem kötelező adatmező kitöltésére – vagy az az előírás, hogy ezzel a mezővel nem is kell foglalkozni. Talán a könyvtárak közti adatcsere, vagy korábbi szabályzatok szerint és/vagy könyvtári rendszerekben történő katalogizálás lehet az oka, hogy nagy ritkán mégis megjelennek ilyen értékek.

1.6 Rekordtípus és dokumentum speciális megjelölése szöveges dokumentumok esetén (Rf/06 vs. 007/01 t)

A rekordtípus (Rf/06) és a szöveges dokumentumok speciális fizikai megjelenéseinek (007/01=t) összehasonlítása is rendkívül kevés abszolút adatot tartalmaz. Azaz a fenti diagram csak a relatív adatok miatt látszik betöltöttnek. Érdeemes megnézni, hogy – például a DNB és a BL esetében – milyen

Rekordtípus és A dokumentum speciális megjelölése (szöveg)

a dokumentum speciális megjelölése (szöveg, 007/01)



6. diagram Rekordtípus és dokumentum speciális megjelölése szöveges dokumentumok esetén

a megoszlás az egyes – kötelezően kitöltendő – rekordtípusok, és milyen a nem kötelező speciális dokumentum megjelölések között szöveges dokumentumok esetén (lásd a 2. és 3. táblázatot).

A szöveg rekordtípus (Rf/06=t) adatai alapján ugyan egyértelmű, hogy a legtöbb dokumentum nyelvi anyag (007/=a). A két német nyelvű könyvtár adatai szerint túlnyomó többségük normál nyomtatás. Az LC-ben alig minősítettek normál nyomtatással (007/01 t=a), ezért ugyanez csak feltételezhető. Ahogy azt se volna érdektelen elemezni, hogy milyen rekordtípusból mennyi a vakírás vagy Bra-

ille-írás. A fenti összehasonlítással azonban ilyen adatok megállapítása nem lehetséges, mert a 007 pozíciót a könyvtárak lényegében nem töltik ki.

Mivel a 007 kitöltése nem kötelező, a végtelenül kevés feldolgozott dokumentum ráadásul vagy ismeretlen, vagy kódolatlan (007/01=t vagy u), kiderül, milyen kevés jelentőséget tulajdonítanak ennek a vezérlő mezőnek.. Különleges eset a LIBRIS, ahol a készlet (Rf/06=o), a NED, ahol sajtóságos módon a Braille-írás (007/01 t=c) és a CSH, ahol a nyelvi anyag egy részét nyomtatottnak, másik részét ismeretlennek minősítették.

2. táblázat A rekordtípus (Rf/06) értékei az LC-ben és a DNB-ben

Kötelezően kitöltendő rekordtípusok megoszlása (rekordfej/06)	LC 13817649 = 100 %		DNB 19026644 = 100 %	
a - nyelvi anyag	12548611	91.0 %	16903888	88.0 %
c - nyomtatott zenemű	136248	1.0 %	578352	3.4 %
d - kéziratos zenemű	8383	~0.1 %	-	-
e - kartográfiai anyag	333929	2.4 %	291148	1.5 %
f - kéziratos kartográfiai anyag	2848	~0.1 %	-	-
g - audiovizuális anyag	120526	0.9 %	10469	~0.1 %
l - nem zenei hangzó anyag	58991	0.4 %	137321	0.7 %
j - zenei hangzó anyag	450795	3.3 %	1008219	5.3 %
k - kétdimenziós (nem kivetíthető) ábrák	127809	~0.1 %	-	-
m - számítógép által kezelt állomány	11017	~0.1 %	-	-
o - vegyes dokumentumcsomag, készlet	19	~0.0 %	20761	~0.1 %
p - vegyes anyag	11844	~0.1 %	-	-
r - háromdimenziós alkotások	46	~0.0 %	-	-
t - kéziratos nyelvi anyag	6583	0.5 %	76433	0.4 %
[érvénytelen, hibás jelzet]	-	-	-	-

3. táblázat A rekordtípus (Rf/06) és a szöveges dokumentumok speciális fizikai megjelenéseinek (007/01=t) metszeteinek rekordszámjai az LC-ben és a DNB-ben

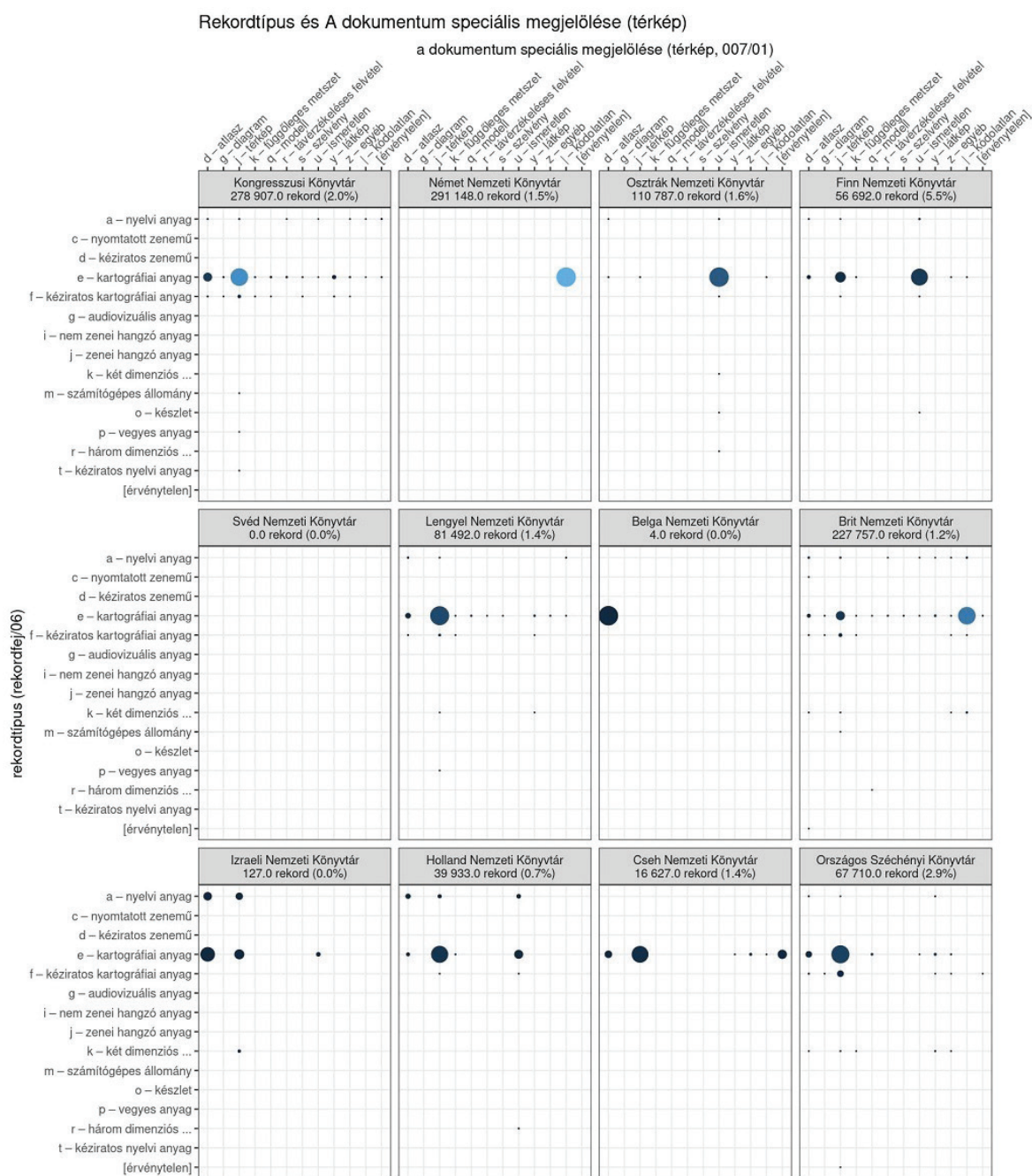
Nem kötelezően kitöltendő dokumentum speciális megjelölése (007/01 t)	LC 13817649 = 100 %		DNB 19026644 = 100 %	
a - normál nyomtatás	8261	0.06 %	10218462	5.37 %
b - nagybetűs nyomtatás	31	~0.00 %	-	-
c - Braille-írás (vakírás)	-	-	-	-
d - szabadlapos kötésben	2	-	-	-
u - ismeretlen	550	~0.00 %	10165227	0.534 %
z - egyéb	44	~0.00 %	-	-
- kódolatlan	60660	0.44 %	-	-
[érvénytelen, hibás jelzet]	14	~0.00 %	1	-

Az OSZK-ban a rekordfej 06 pozíciója esetében egyáltalán nem minősítik a szöveges dokumentumok speciális megjelölését. Nyilván ama sablon szerint, melyet a katalogizáló rendszerük tartalmaz.

Az ember nem gondolná, hogy mind a DNB-ben, mind a BL-ben több forrást dolgoztak föl, mint az LC-ben! BL = 19,154.445 rekord, DNB = 19,026.644 rekord, LC = 13,817.649 rekord. Az angolban (BL) azonban a nem kötelező minősítések (00X) megadása kevésbé részletes, mint a másik kettőben.

A két német nyelvű könyvtárban (DNB, ÖNB) valószínűleg szabályzat rögzíti, hogy a dokumentum speciális megjelölése esetén alapértelmezés a z=egyéb, illetve az u=ismeretlen. Az összes többi könyvtárban a 2. és 3. táblázatban, illetve diagramokban látható adatok az egészhez képest elenyészőek, és véletlenszerűen keletkezettek. Vajon ezekben szabályzat írja ezt elő?

Mivel a 65X-es tárgyszó mezőkben például a vakírást, sőt akár annak speciális fajtáit is osztályoz-



7. diagram Rekordtípus és dokumentum speciális megjelölése térképek esetén (Rf/06 vs. 007/01 a)

hatják, összetett kereséssel a nyelvi anyagnak megfelelő tárgyszóval, és a vakírás vagy a Braille-írás tárgyszóval keresve az ilyen adat mindezek ellenére megtalálható – csak éppen ezt a fajta keresést legfeljebb a nagyon gyakorlott referenz könyvtáros képes megvalósítani a mai online katalógusokban, lévén, hogy a felhasználók többsége összetett keresésre csak nagyon ritkán fanyalodik.

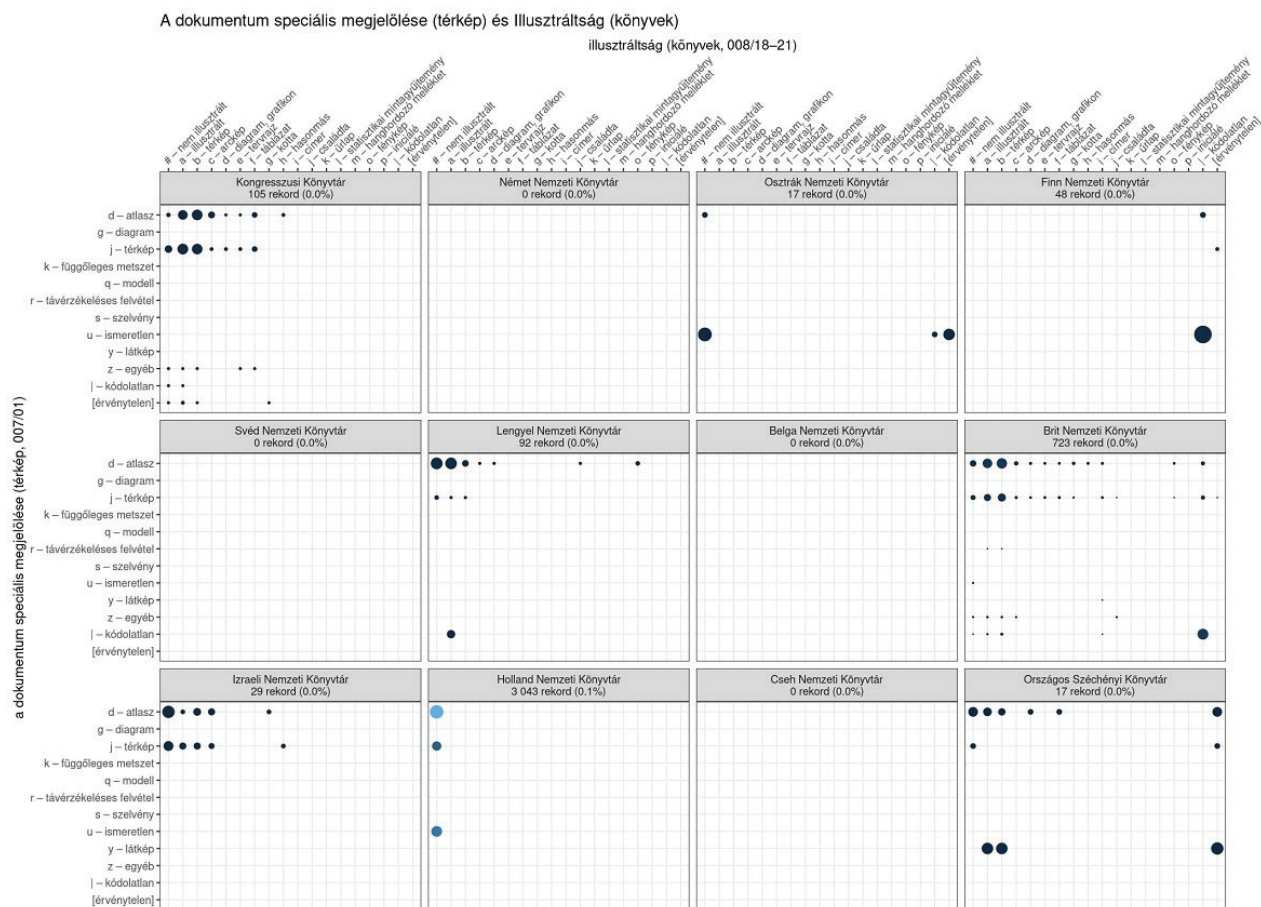
A 007 és 008 mezők kitöltetlenségének alapvetően az az oka, hogy a dokumentumtípusok speciális osztályozása – ahogy erre előzőleg már utaltunk – formai tárgyszavakkal/deszkríptorokkal (655) sokkal rugalmasabban megoldható. Ráadásul a fizikai jellemzők is könnyen megadhatók szövegesen is az 588 mezőben (Megjegyzés a terjedeletről/fizikai jellemzőkről).

A 300-as szöveges mező \$a almezőjében ugyancsak megadható a fizikai egységek száma, a fizikai jellemzők. Az 515-ös szöveges mezőben pedig az elektronikus dokumentumok típusa.

Mindez nem változtat azon, hogy egy, a jelenleginél sokkal jobban kidolgozott egyszerű tipológiai és egyéb jellemzők osztályozási rendszere a 007-es 008-as vezérlőmezőkben kifejezetten hasznos lehetne – amennyiben a katalógusok keresőit ezek könnyű, felhasználóbarát alkalmazására felkészítenék.

1.7 Rekordtípus és dokumentum speciális megjelölése térképek esetén (Rf/06 vs. 007/01 a)

A térképek esetében még kevésbé használják a 007 vezérlő mező 01 pozícióját osztályozásra. A két német nyelvű könyvtárban, az NFI-ben és a BL-ben minden kartográfiai anyagot a kódolatlan vagy az ismeretlen minősítéssel látták el, lehet hogy ez az alapértelmezett besorolás. Az LC-ben ugyan látható a kartográfiai rekordok sűrűsödése atlaszok és térképek esetén (007/01=d, ill. =k), de az előforduló kevés adat legfeljebb véletlenszerű, nem lehet belőlük következtetéseket levonni. Itt még az se feltételezhető, hogy a nemzeti könyvtárban alapértelmezéseket adnának meg a kitöltési szabályzatok.



8. diagram Könyveknek minősülő térképek és illusztráltságuk

A minősített kartográfiai anyagok állománya a teljeshez viszonyítva csak az NFI-ben számottevőbb (5.5%), utána az OSZK (2.9%), majd az LLC következik (2%). A kartográfiai anyagok térképészeti minősítése, például a látképeké (007/01=y), egyébként az OSZK-ban a legváltozatosabb, míg az LC, a BNPL és a BL ettől némiképp lemaradva hasonlóan változatos.

A pozíciók gyenge kitöltöttsége összefügg azzal, hogy a 300-as Terjedelem/Fizikai jellemzők mező \$a és \$b almezői, továbbá az 588-as Megjegyzés a terjedelemtől/fizikai jellemzőkről megjegyzés mezőkben ugyancsak minden térképtípust meg lehet adni szövegesen (lásd a 4. táblázatot).

1.8 Könyveknek minősülő térképek és illusztráltságuk

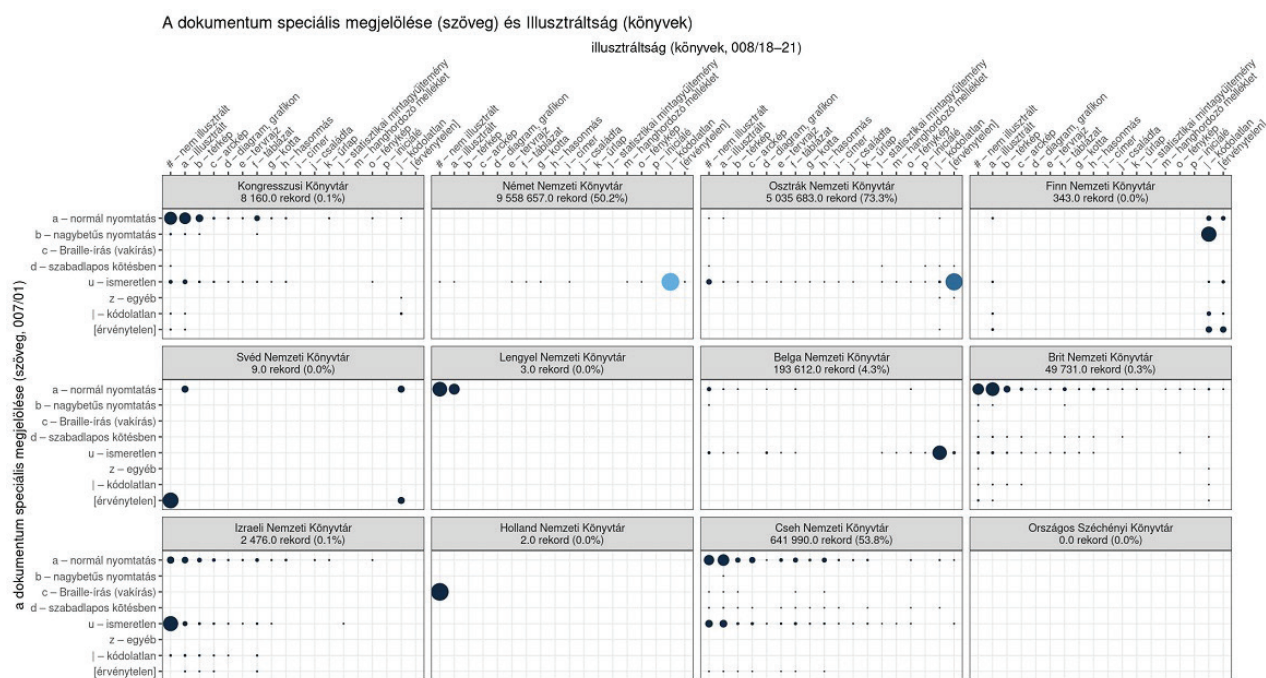
A rendkívül kevés adat, illetve az adatok teljes hiánya ugyanarra enged következtetni, mint az előző, 8. diagramcsoport esetén. Minden jel szerint a kartográfiai anyagok illusztráltságának finomabb osztályozását a 008 mező 18–21. pozíciójában nem tekintik fontosnak. Ez talán azért van, mert – hasonlóan a térképek speciális osztályozásához az előbbi diagramcsoport tanulságához – a térképek speciális illusztráltságát részletesen le lehet írni szövegesen az 588-as Megjegyzés a terjedelem-

ről/fizikai jellemzőkről mezőben (lásd a 4. táblázatot), a 65X tárgyiszó mezőkben pedig pontos, speciális, természetes nyelvű tárgyiszóval/deszkríptorokkal is osztályozni lehet.

Az LC-ben és BL-ben (némileg a BNPL-ben is) a kevés feldolgozott rekord ugyanakkor arra utal, hogy a térképeknek nagyon is van illusztráltsága. A NED-ben szerepel a legtöbb minősítése a kartográfiai anyagoknak (3043 db), de ezeket egységesen a „nem illusztrált” (008/18–21=#) minősítéssel látták el. Az OSZK 17 darab, illusztráltság szempontjából változatosan minősített kartográfiai anyaga se értékelhető.

1.9 Szöveges könyvek illusztráltsága (007/01 t vs. 008/18–21)

A szöveges dokumentumok illusztráltságát egyes könyvtárakban a térképekénél nagyobb számban minősítik, de a DNB-ben, az ÖNB-ben és az NFI-ben a kódolatlanság, ill. az érvénytelenség alapértelmezésével. Más könyvtárakban olyan kis rekordmennyiségekről van szó, hogy minősítésük nem is értelmezhető. Az OSZK-ban a katalogizáló rendszer úgy definiálták, hogy automatikusan nem minősít semmit a 007-es mezőben, és ma már ennek oka nem ismert.



9. diagram Szöveges könyvek illusztráltsága

4. táblázat A rekordfej és a 00X vezérlő mezők dokumentumtípusok osztályozására való pozíciói és a tartalmilag velük összefüggő, ugyancsak dokumentumtípusok leírására használható szöveges megjegyzésmezők

A rekordfejben és a 00X vezérlőmezőkben	A szöveges megjegyzésmezőkben
Rekordfej (rekordtípus 06 c=nyomtatott zenemű; d=kézirat zenemű)	254 \$a A kotta típusának megevezése
Rekordfej (rekordtípus 06 m=számítógép által kezelt állomány; 007/00 dokumentumkategória c=elektronikus dokumentum)	256 \$a és \$b Elektronikus dokumentumok jellemzői
007/01 a dokumentum speciális megjelölése: térkép 007/01 a dokumentum speciális megjelölése: szöveg 008/18–21 könyvek: illusztráltság	300 Terjedelem, fizikai jellemzők \$a A dokumentum fizikai hordozójának a fajtája \$b Egyéb fizikai jellemzők, illusztrációk 588 Megjegyzés a terjedelemtől/fizikai jellemzőkről
008/24–27 tartalmi jellemzők: könyvek	513 Megjegyzés a kutatási jelentés típusáról és idejéről \$a A jelentés típusa.
Rekordfej (rekordtípus 06 m=számítógép által kezelt állomány; 007/00 dokumentumkategória c=elektronikus dokumentum)	516 Megjegyzés az elektronikus dokumentum típusáról vagy adatairól

Figyelemre méltó, hogy a szöveges dokumentumok (007/01=t) illusztráltságának (008/218–21=a) vagy semmi jelentőséget nem tulajdonítanak, vagy a 008-as mezőben nem kezdenek vele semmit. A német könyvtárakban (DNB, ÖNB) különös módon legfeljebb annyit tesznek, hogy a rekordok felében-háromnegyedében a 007/01 szöveget u=ismeretlennek, a 008/18–21 pozíciókon pedig az illusztráltságot kódolatlanak vagy érvénytelennek minősítik.

Több nemzeti könyvtárban (LC, DNB, ÖNB, KBR, BL) elenyésző számú esetben tapasztalható volt, hogy igenis lenne itt osztályozni való, de aztán mintha gyorsan föladták volna.

Mind az előző 9. mind erre a diagramcsoportra érvényes, hogy a *MARC21 az illusztrációk osztályozásának dolgában különösen redundáns*. A 300-as szöveges mező \$b almezőjében (Egyéb fizikai jellemzők) ugyancsak megadhatók az illusztrációk. *Ez önmagában nem baj, ha a MARC21-en szabályoznák egymáshoz viszonyítva a két mező tartalmának használatát*. Ez egyébként az előző diagramcsoportokban említett redundanciák esetében sem történik meg.

1.10 Szöveges dokumentumok tartalmi jellemzői (007/01 t vs. 008/24–27)

A szöveges dokumentumok (007/0=t) tartalmi jellemzőinek (008/24–27) megadására is ugyanaz jellemző, mint az előző, 9. diagramcsoportra. Alig van a könyvtárakban feldolgozás, a németben csak csupa kódolatlan (008/24–27=l), az ÖNB-ben meg csupa ismeretlen.

Csak a DNB-ben (50.2%), az ÖNB-ben (73.3%) és a CSH-ban (53.8%) nagy a minősítések száma, de döntő többségük alapértelmezésnek látszik. Az OSZK-ban a szöveges dokumentumok tartalmi jellemzőit egyáltalán nem minősítik.

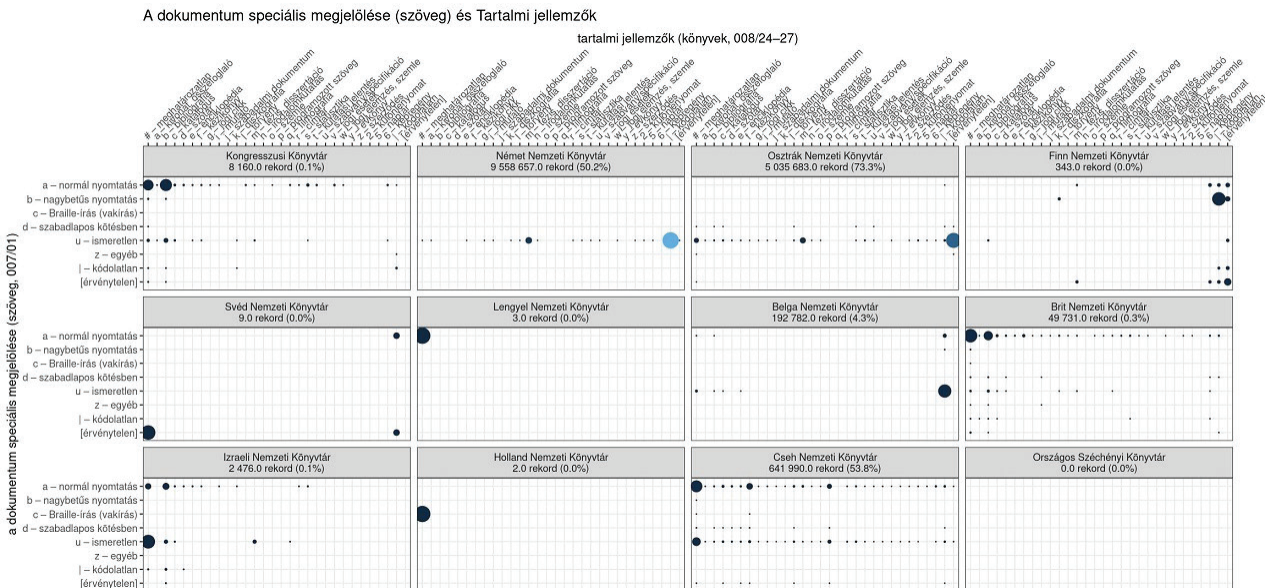
Itt is lehetséges, hogy a tartalmi jellemzők szerinti dokumentumosztályozást a 655 formai tárgyszavak mezőjében oldják meg, és ezért nem foglalkoznak a 008/14–27. pozíciók kitöltésével. Sajnálatos, hogy e téren is hiányos a szakmai nyilvánosság.

A kutatási jelentéseket egyébként részletesen osztályozni lehet az 513-as mező \$a almezőjében (lásd a 4. táblázatot). A szöveges mezők használatára vonatkozóan már a korábbi diagramcsoportokban jeleztük, hogy kereshetőségük egyáltalán nem megoldott.

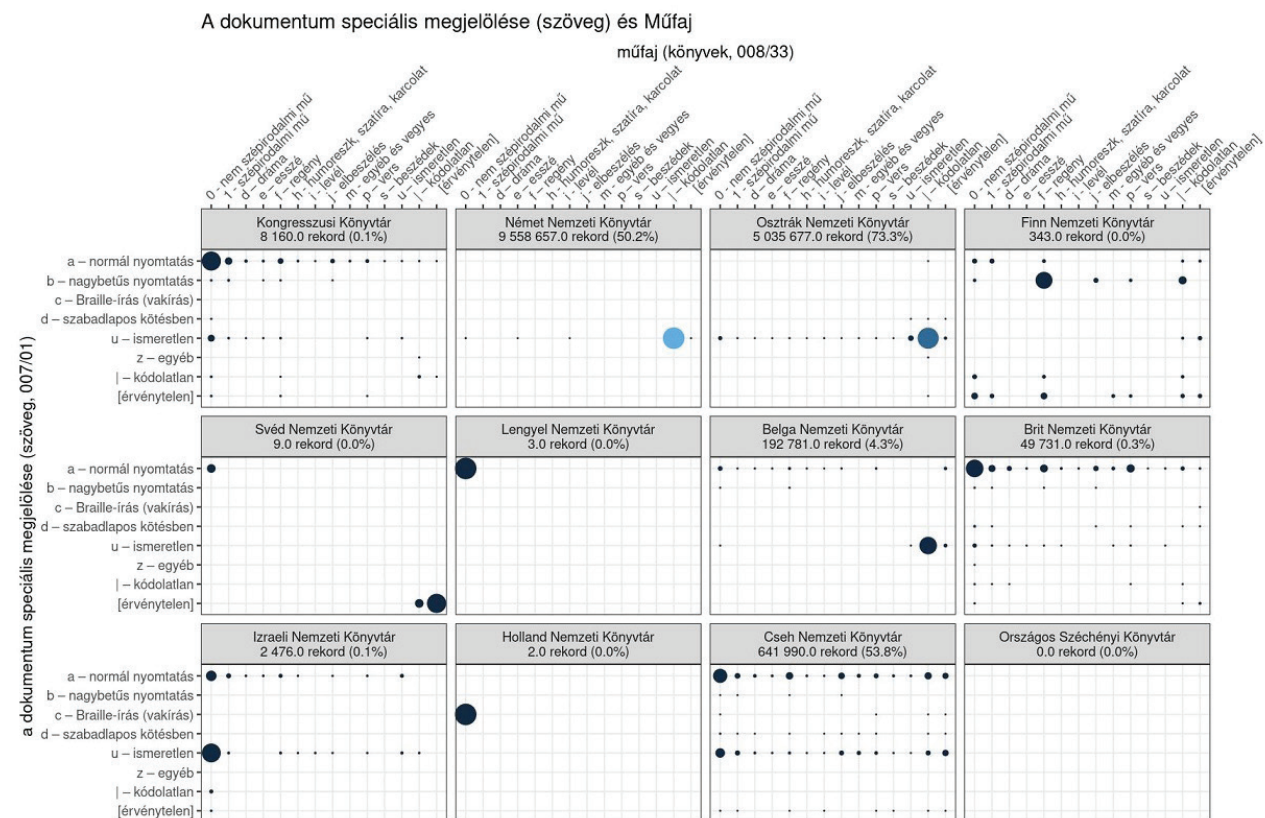
Megint csak figyelemre méltó az ÖNB gyakorlata, ahol a dokumentum speciális megjelölésére az ismeretlen minősítést (007/00=u), viszont a könyvek tartalmi jellemzőjeként, ha ismeretlennek minősítik azokat, a kettős keresztet (#) adják meg a köz helyett.¹⁶

1.11 Szöveges könyvek műfaji jellemzői (007/01 t vs. 008/33)

A szöveges dokumentumok (007/01=t) műfajának megadására is az előbbi 11. diagramcsoport jellemzői érvényesek. Alig van a könyvtárakban feldolgozás, a DNB-ben és az ÖNB-ben meg csupa kódolatlan (008/33=l). Egyedül a CSH-ban van érdemleges műfaji minősítése a szöveges dokumentumoknak. A többi könyvtárban – ahhoz képest, milyen nagy a szöveges dokumentumok állománya – véletlen-



10. diagram Szöveges dokumentumok tartalmi jellemzői



11. diagram Szöveges könyvek műfaji jellemzői

szerű, hogy esetenként mikor minősítettek műfaj szerint. Az OSZK-ban a szépirodalmi dokumentumok műfaji jellemzőit – hasonlóan az előző fejezetben tárgyalt 24–27. pozíciójának értékeihez – egyáltalán nem minősítik.

Itt is fölvetődik, hogy a műfajok szerinti dokumentumosztályozást valószínűleg a 655 formai tárgyszavak mezőjében oldják meg, és ezért nem foglalkoznak az egyébként sem kötelező 008/33. pozíciók kitöltésével. Kétségtelen ugyanis, hogy

a dokumentumok formai osztályozására nagyszámúrendileg sokkal, de sokkal több tárgyszóra van szükség, semmint, hogy azokat a vezérlőmezők pozícióiba egykarakteres jelzetekkel bele lehetne préselni.

Megnézve a DNB kódolatlan minősítésű rekordjait, olyanok vannak benne, mint pl. életrajz, feladatgyűjtemény, gépkönyv stb., és tényleg, se a 008/33, se a 008/25–27 nem rendelkezik ezekhez megfelelő tartalmi jellemzőkkel, illetve műfajokkal. Ugyanakkor olyanok is szerepelnek a kódolatlanok között, mint például regények, esszék, melyeknek van a 008/33-ban típusuk, az „f=regény”, „e=esszé”. Az életrajzok pedig az „1=szépírodalom” típusba sorolhatók.

Minden jel szerint a könyvtárak nem akarnak se az illusztrációk, se a tartalmi jellemzők, se a műfajok besorolásaival a vezérlő mezőkben (00X) vesződni. Feltehetően azért, mert a szűkös, néhány pozícióhoz kötött, egykarakteres osztályozási választék használata sok fejtöréssel jár, és a

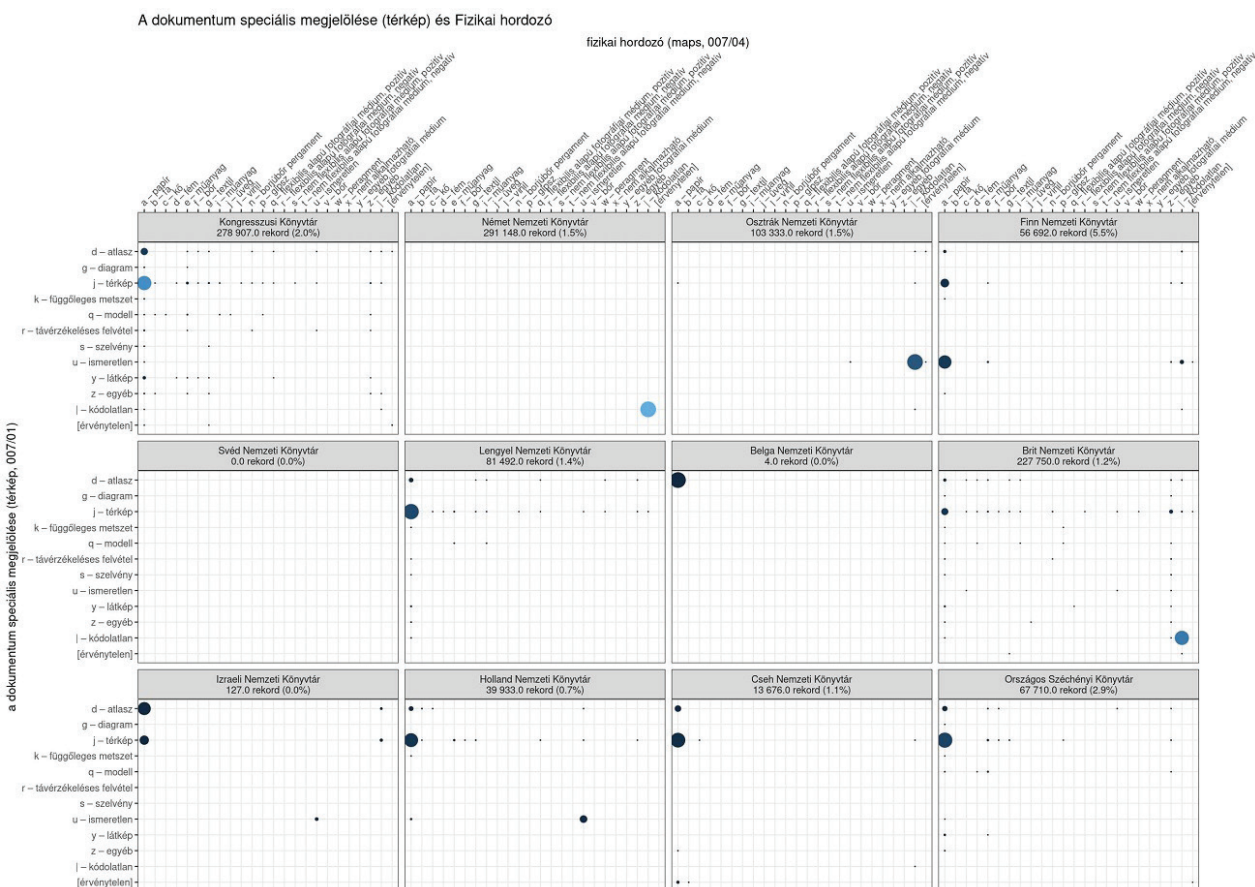
vonatkozó MARC 00X mezők értékei se nem elégségek, se nem alkotnak koherens műfaji-tartalmi rendszert.

Mindez a 008 könyvek esetére vonatkozik, de valószínűleg ugyanez a helyzet az összehasonlításba nem bevont időszaki kiadványok és zenei anyagok esetén is, ahol szintén vannak ilyen tartalmi jellemzők és műfajok.

1.12 A térképajták (007/01) fizikai hordozójuk (007/04)

A diagramcsoport tipikusan példázza a 007-es mező nem kötelező voltának a következményeit.

A mező 04. pozícióján a mindenkori dokumentumtípus (esetünkben a térképek) hordozóit kellene osztályozni. Ez vitathatatlanul fontos keresési szempont lehet. Ennek ellenére a kitöltöttsége minimális. Akárcsak az előző fejezetben, ez már önmagában fölveti, hogy akkor egyáltalán miért van a viszonylag elenyésző esetekben mégis kitöltve a pozíció?



12. diagram Térképajták fizikai hordozói

A legtöbb könyvtár a kevés kitöltést eleve elintézi azzal, hogy vagy a kódolatlan (007/04=l), vagy a papír (007/04=a) értéket adja meg.

Ugyanakkor az LC, a BNPL és a BL esetében még a viszonylag kevés esetében is jól látható már, hogy nagyon sok más hordozótípus is előfordulhat, nemcsak a papír. Ez vonatkozik az OSZK-ra is. Teljes kitöltöttség esetén a diagramok feltehetően nagyon látványos képet mutatnának.

Kérdés, mennyire osztályozzák a forrásokat a hordozó szempontjából más mezőkben (főleg a 65X mezőkben)?

1.13 Az ugyancsak dokumentumtípusokhoz használható más mezők

Redundáns mezők a 00X mező pozícióinak tartalmával (lásd a 2. táblázatot).

2. A 00X adatmezők „dicsérete”

A 00X kódolt fizikai és meghatározott jellemzők („Fixed-Length Data Elements - Additional Material Characteristics) a kereshetőség változatosságának lehetőségei. Általuk magától értetődő, átfogó tartalmi tájékozódás lehetséges. Főképp a keresés elején játszhatnának hasznos szerepet: a végfelhasználó számára kiderülhetne, milyen sokféle irányban indulhatna el, a különféle dokumentum-

és tartalomtípusok szerint. Nagyon egyszerű informatikai megoldásokkal lehet(ne) közelebb hozni a laikus keresőhöz ezeket a különféle szempontok szerinti adottságokat. Ezek ma, talán az egy Kongresszusi Könyvtárat kivéve még a nagy nemzeti könyvtárakban sincsenek igazán kihasználva.

Az LC keresőoldalán¹⁷ a keresőszótól balra például nagyon egyszerű módon egy legördíthető mezőben többek között dokumentumtípusok szerint lehet tovább szűkíteni a keresést (2. ábra). A felhasználó egy pillantással áttekintheti, a megadott tárgy esetén melyikből hány rekord létezik.

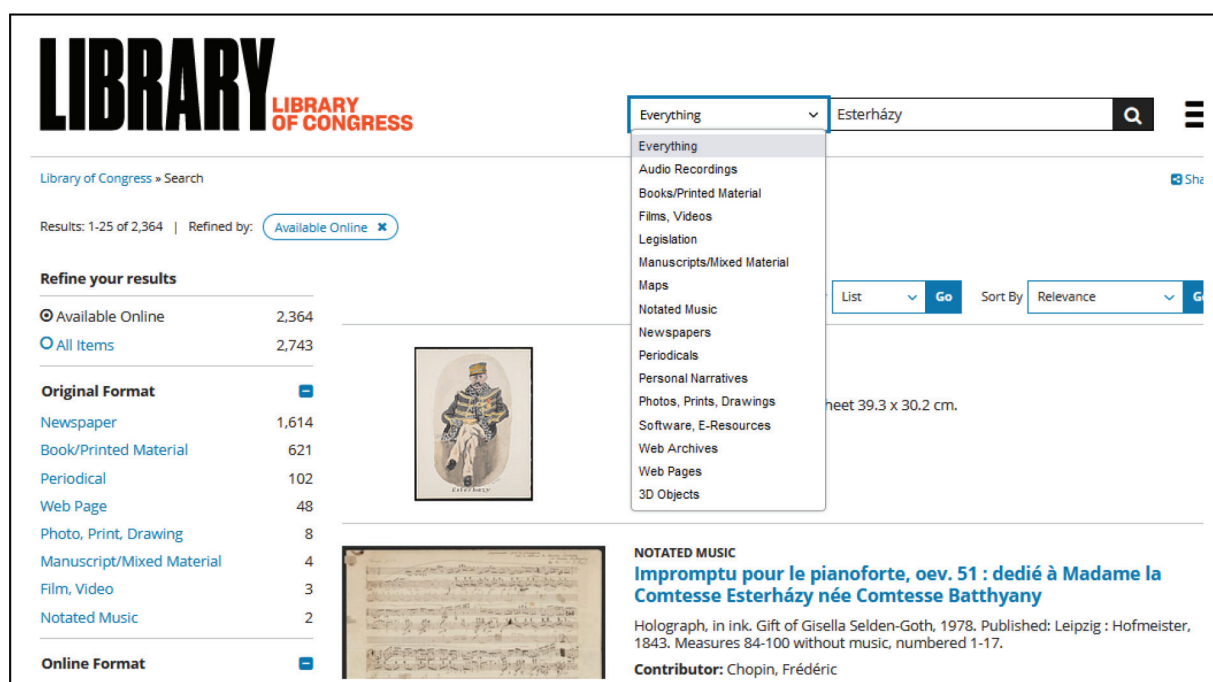
A 2. ábrán bemutatott oldal folytatása, a bal szélen további 00X mezők szerinti kereshetőségekkel, mint például a dátumok (3. ábra).

3. Tárgyszómezők (tárgyi melléktételek)





3.1 Személynév tárgyszó

A tárgyyszómezők a tartalom, forma (dokumentumtípus), idő, földrajzi elhelyezkedés szerinti legfontosabb mezőcsoportot alkotják. Az 5. táblázatban a könnyű áttekinthetőség érdekében csak a legfontosabbnak vagy legjellemzőbbnek ítélt almezőket tüntettük föl.

Feltűnő, mennyivel több személynévvel tárgyszavazzák a tartalmat a LCL-ben (11.79%), BNPL-ben (18.51%) és NIZ-ben (14.18%) (félkövéren jelöltük),



2. ábra Legördíthető választék a Kongresszusi Könyvtár (LC) keresőfelületén¹⁸

Original Format			1 print : pochoir on paper ; sheet 39.3 x 30.2 cm. Contributor: Shahn, Ben Date: 1984
Newspaper	1,614		
Book/Printed Material	621		
Periodical	102		
Web Page	48		
Photo, Print, Drawing	8		
Manuscript/Mixed Material	4		
Film, Video	3		
Notated Music	2		
Online Format			NOTATED MUSIC Impromptu pour le pianoforte, oev. 51 : dédié à Madame la Comtesse Esterházy née Comtesse Batthyany Holograph, in ink. Gift of Gisella Selden-Goth, 1978. Published: Leipzig : Hofmeister, 1843. Measures 84-100 without music, numbered 1-17. Contributor: Chopin, Frédéric Date: 1842 View with Search Result: View 4 Images Resource: View All Images PDF
Image	2,314		
Online Text	1,972		
PDF	1,929		
Web Page	7		
Video	3		
Compressed Data	2		
Date			PHOTO, PRINT, DRAWING [Trademark registration by J. B. Weaver & H. Schandain for Garwood's Esterhazy brand Perfumery] 1 item ; sheet 26 x 39 cm Date: 1885
2000 to 2099	19		
1900 to 1999	678		
1800 to 1899	1,507		
1700 to 1799	4		PHOTO, PRINT, DRAWING Eisenstadt, der Schlossplatz in umgestaltung

3. ábra További 00X mezőválaszték a Kongresszusi Könyvtár (LC) keresőfelületén

5. táblázat A személynév tárgyszó legjellemzőbbnek ítélt almező-értékei a teljes állomány százalékában

600 személynév tárgyszó	LC	DNB	ÖNB	NFI	LIBRIS	BNPL	KBR	BL	NIZ	NED	CSU	OSZK
\$a rendszó	11.79	2.28	0.25	4.88	8.37	18.51	0.02	5.78	14.18	8.36	5.10	6.62
\$b uralk. sorszám	0.25	0.05	0.00	0.06	0.21	0.71	0.00	0.14	0.36	0.38	0.15	0.42
\$c méltós., fogl.	1.47	0.21	0.00	0.65	0.85	2.39	0.01	0.94	2.13	1.74	0.64	0.66
\$d kronolog. kieg.	8.85	1.85	0.03	2.16	5.68	16.04	0.02	4.00	11.18	7.39	4.72	6.16
\$e kapcsolatjelölő	0.01	-		0.01	0.00	0.00	0.00	0.01.	-	-	-	
\$g álland. mellékn.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	-	0.22
600j egyéni név.	-	-	-	-	0.00	-	0.00	0.00	-	-	-	0.22
600m hangszer ¹⁹	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	-	0.00	0.01
600q teljes név	0.61	-	0.00	0.00	0.08	0.00	0.00	0.29	0.47	-	0.04	-
600t címreláció	0.86	0.19	0.00	0.16	0.50	3.88	0.00	0.44	1.36	-	0.15	0.32
6000 bes. rek. azon.	0.15	1.99	0.00	3.56	3.76	0.00	0.00	0.26	0.00	-	-	-
6001 URI	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	-	-
6002 forrás	0.18	2.28	0.00	0.00	0.25	18.48	0.00	0.28	14.18	-	5.04	-

rádásul az utóbbi kettő nem is tartozik a legnagyobb nemzeti könyvtárak közé. Az OSZK a középmezőnyben van (6.62%).

Az uralkodói nevek tárgyszóként való gyakori használata (LC, BNPL, NIZ, NED és OSZK) összefügghet a történelmi művek nagyobb arányával is.

A kronologikus kiegészítő gyakoribb használata (BNL: 16.04%, NIZ: 11.18%, LC: 8.85%, NED: 7.39%) a személynév tárgyszó gondos szerkesztésére utal.

Erre utal, hogy ahol ez az érték nagyobb, ott a \$c foglalkozás kiegészítő értéke is nagyobb. E téren az OSZK a középmezőnyben foglal helyet (6.16%).

Feltűnő, hogy a tárgyszóként használt név és a mű kapcsolatát leíró, szerepet, közreműködést jelölő \$e kapcsolatjelölő használata milyen elenyésző, számos könyvtárban egyáltalán nem használják (az OSZK által használt HUNMARC eleve nem tartalmazza). Feltehetően azért, mert valójában

redundáns, a cím és szerzőségi közlésben amúgy is szerepel, ráadásul ugyanaz a név több szerepet, közreműködést is betölthet, és akkor az online mutatóban annyiszor is fog szerepelni (hacsak – gondosan – a katalogizáló rendszert úgy nem készítették, hogy ezt az almezőt ne vegye figyelembe).

A \$j az OSZK-ban a HUNMARC szerint az utónevet tartalmazza, feltűnően nagy értékét ez magyarázza; a MARC21 szerint a \$j a bizonytalan, redundáns nevet pontosító (viszonyminősítő) adat és jóformán nem használják.²⁰ A \$m a HUNMARC szerint rokon-sági kiegészítő, a MARC21 szerint a mű hangszere.

A \$q teljes név inkább az angol nyelvű (LC, BL) és az őket követő izraeli (NIZ) könyvtárakban használatos, ahol inkább felveszik önállóan a rövidített nevet.

Fontos jelenség, hogy újabban bekerült a MARC21-be a besorolási adatok²¹ \$0 azonosítója, melyet a legtöbb könyvtárban az új rekordokban mindig megadnak, egyedül a NED, CSU és az OSZK nem alkalmazza. Közéltően ez vonatkozik a \$1 fizikai (három dimenziós) tárgyak (hagyományosan nem bibliográfiai anyagok) azonosítójára is (real world object, URI). Az utóbbi azért is jelentős, mert összefügg a mesterséges intelligencia hamarosan robbanásszerű terjedésével, melyben az ilyen tárgyakat azonosítva kezelik. Mindkettő viszonylag új fejlemény a MARC21-ben, nem is biztos, hogy minden könyvtár katalóguskezelő rendszerében újként definiálhatók, ha a rendszert úgy telepítették, hogy abba minden eleve „bele lett égetve”, és a felhasználói részről nem lehet rajtuk módosítani. Az OSZK (együtt a NED-del és a CSU-val) egyáltalán nem használja, mivel a HUNMARC nem tartalmazza. Ez előbb-utóbb komoly lemaradással jár.

Egyes könyvtárakban nagy gondot fordítanak arra, hogy minden személynév tárgyszóhoz megadják a forrást (DNB, BNPL, NIZ, CSU). Mivel ez a HUNMARC-ból a személynevek dolgában hiányzik. Az OSZK-ban a személynevekhez nem adnak meg forrást (a többihez igen). Érdekes, hogy az LC

is rendkívül szegényes ennek az almezőnek a használatában. Ez az állomány számítógépes elemzésének a lehetőségét csökkenti, mivel nem lehet kapcsolatot találni az alkalmazott tárgyszórendszerek, teauruszok, névterek és a katalogizált rekordok között.

Feltűnő, hogy az ÖNB-ben és a KBR-ben a személynév-tárgyszavak (továbbá a szaktárgyszavak és földrajzi tárgyszavak) használata minimális. Viszont mindkettőben jelentős a szabadon választott tárgyszavak és a formai tárgyszavak használata.

A testületi nevek és rendezvénynevek dolgában a fentiekhez hasonló a helyzet, csak a rekordoknak kisebb százalékához kapcsolódnak.

3.2 Kronologikus tárgyszó

A BNPL (39.91%), CSU (10.45%), NED (6.85%) és NFI (2.91%) a nemzeti könyvtárakban a kronologikus tárgyszó használata átlagon felüli. A könyvtárak többségében a forrásadatot is gondosan megadják.

Egyedül az OSZK és az LC nem minősíti külön a kronologikus tárgyszavakat, ami veszteség. Az OSZK azért, mert az AMICUS katalóguskezelő rendszerében ilyet nem definiáltak, holott a Köztársaságban, melyet az OSZK-ban is használnak a tárgyszavak kiválasztására, ezek a tárgyszavak minősítve vannak.

3.3 Szaktárgyszó

Nagyon feltűnik, hogy a nemzeti könyvtárak többségéhez képest, ahol a rekordok jóval több, mint a fele, ill. harmada kap a tartalmát meghatározó 650-es szaktárgyszót, az OSZK-ban ez még a rekordok harmadát se éri el (26.16%). A 12 könyvtár közül a középmezőnyben van (a maga 26.16%-ával a 9.), holott az OSZK-ban hagyományosan színvonalas a tartalomfeltáró munka. Ennek oka, hogy csak 2004-től kezdve adnak meg tárgyszavakat. Az addig leírt állományt visszamenőleg nem látták el tárgyszavakkal, de ha retrospektív feldolgozásra kerül sor, akkor ezt pótolják.

6. táblázat A kronologikus tárgyszó legjellemzőbbnek ítélt almező-értékei a teljes állomány százalékában

648 kronologikus tárgyszó	LC	DNB	ÖNB	NFI	LIBRIS	BNPL	KBR	BL	NIZ	NED	CSU	OSZK
\$a rendszó	-	2.31	0.35	2.91	0.74	39.91	0.00	0.41	0.02	6.85	13.49	-
\$2 forrás	-	2.31	0.04	2.91	0.53	39.85	0.00	0.41	0.00	6.84	12.12	-

7. táblázat A szaktárgyszó legjellemzőbbnek ítélt almező-értékei a teljes állomány százalékában

650 szaktárgyszó	LC	DNB	ÖNB	NFI	LIBRIS	BNPL	KBR	BL	NIZ	NED	CSU	OSZK
\$a rendszó	71.91	18.98	0.41	39.48	42.5	69.84	12.28	35.68	66.76	56.41	46.14	26.16
\$b földr. névelem	0.01	-	0.01	0.00	0.01	0.00	0.23	0.00	0.00	0.02	-	-
\$c esemény helye	0.00	-	0.00	-	0.01	0.00	0.00	0.00	0.00	0.01	-	0.05
\$0 besor. rek. azon.	0.14	14.95	0.01	39.49	13.94	0.00	0.00	2.83	0.02	-	-	-
\$2 forrás	0.18	18.98	0.02	39.48	28.00	69.73	4.79	4.25	66.64	49.49	46.10	26.16
\$6 mezőkapcsoló	0.73	-	-	-	0.01	0.00	-	0.02	0.00	-	-	-

8. táblázat A földrajzi név tárgyszó legjellemzőbbnek ítélt almező-értékei a teljes állomány százalékában

651 földrajzi tárgyszó	LC	DNB	ÖNB	NFI	LIBRIS	BNPL	KBR	BL	NIZ	NED	CSU	OSZK
\$a rendszó	23.27	5.86	0.18	16.02	8.57	30.16	0.14	9.85	15.23	13.03	13.61	12.27
\$b egyéb földr. elem	0.00	-	-	-	-	0.00	-	0.00	-	-	-	0.14
\$0 besor. rek. azon.	4.69	5.17	0.07	16.02	2.15	0.00	0.00	1.46	0.00	-	-	-
\$2 forrás	4.75	5.86	0.08	16.02	3.30	30.10	0.00	2.78	15.18	10.04	13.57	0.00

Külön ki kell emelni, hogy az OSZK azon kevés nemzeti könyvtár egyike, amely a már említett nyilvánosan elérhető (jelszóval védett) átfogó tezaurusszal (Köztaurusz) rendelkezik.

Még több könyvtárra érvényes, hogy igyekeznek a legtöbb esetben megadni a szaktárgyszó \$2 forrását és az új rekordok esetén a \$0 besorolási rekordazonosítót. A szaktárgyszavakhoz az OSZK-ban mindig megadják a forrást, a besorolási rekordazonosító almezője pedig a HUNMARC-ban nem szerepel.

Számos almezőt csak ritkán töltenek ki, ami a többi tárgyszó esetében is így van. Ez azonban nem jelenti azt, hogy ezek feleslegesek. A speciális kereséskor jól használhatók – lennének, ha a könyvtári rendszereket erre felhasználóbarát formában felkészítenék.

3.4 Földrajzi név tárgyszó

Földrajzi nevek dolgában a szaktárgyszavakhoz hasonló a helyzet: vannak könyvtárak, melyekben aránylag gyakoriak, és a forrást is pontosan megadják. Érdekes, hogy a szaktárgyszavakkal (650) és dokumentumtípusokkal (655) szemben a földrajzi nevekhez az OSZK-ban nem adják meg a forrást, pedig ugyanúgy, mint az említettek, ezek a tárgyszavak is kötelezően a Köztauruszból származnak. Ha új tárgyszó megadására kerül sor, ezeket mindig föl kell venni a Köztauruszba is.

Az OSZK ugyanúgy a 9. helyen van (12.27%), és alig van olyan rekord, melyben megadják a földrajzi név forrását, holott ezeket is a Köztauruszból választják.

3.5 Szabadon választott tárgyszó

Feltűnő, hogy a DNB, ÖNB, LIBRIS, KBR és BL katalógusában milyen nagy arányban használnak szabadon választott tárgyszavakat, ami némileg magyarázza, hogy ezekben a könyvtárakban a többi tárgyszótípus használata relatíve gyengébb. Az még feltűnőbb azonban, hogy mennyire ritkán fordul elő, hogy megadják a \$2 forrást, holott gyakori, hogy azt valahonnan szintén választják, de nem információkereső nyelvi szótárból vagy osztályozási rendszerből. Elvileg még azoknak a szabadon választott tárgyszavaknak is van „forrása”, melyeket a dokumentum szavai közül választanak (ezek a kulcsszavak, forrásnevük a „kulcsszó”), vagy amelyeket a feldolgozó emlékezetből választ (akkor a forrás a „feldolgozó” vagy az annak megfelelő kódszó).

3.6 Formai tárgyszó

A CSU (68.63%) és a BNPL (57.41%) gyakorlatában feltűnően sok esetben megadják a dokumentum típusát. Más könyvtárakban is feltűnik a gyakoriság. Ez természetesen következik abból, hogy minden dokumentum valamilyen típushoz tartozik.

9. táblázat A szabadon választott tárgyszó legjellemzőbbnek ítélt almező-értékei a teljes állomány százalékában

653 szabadon választott tárgyszó	LC	DNB	ÖNB	NFI	LIBRIS	BNPL	KBR	BL	NIZ	NED	CSU	OSZK
\$a rendszó	1.47	32.14	12.76	0.71	12.40	0.00	22.38	14.38	0.74	0.12	2.62	0.05
\$2 forrás	0.00	-	-	0.00	0.00	0.00	-	0.00	-	-	-	-

10. táblázat A földrajzi név tárgyszó legjellemzőbbnek ítélt almezőértékei a teljes állomány százalékában

655 formai tárgyszó	LC	DNB	ÖNB	NFI	LIBRIS	BNPL	KBR	BL	NIZ	NED	CSU	OSZK
\$a rendszó.	11.04	26.80	20.47	6.36	8.64	57.41	0.98	6.87	11.17	17.32	68.63	25.92
\$0 besor. rek. azon.	6.27	26.74	16.67	6.36	6.77	0.00	0.00	2.30	0.00	-	-	-
\$2 forrás	10.68	26.80	18.78	6.36	6.06	57.30	0.72	4.42	11.15	17.08	68.59	25.92

Kétségtelen, hogy ezek közül csak a speciálisabb típusokat adják meg.

Ebből a szempontból az OSZK is jól áll. Ráadásul a formai tárgyszavak esetén is mindig gondosan rögzítették, hogy ezek az ugyancsak az OSZK által kezelt dokumentumtípus-tezauruszból származnak.²²

3.7 Áttekintő diagram

Jól látható a 650-es szaktárgyszók dominanciája, de kiemelkednek a földrajzi név, a szabadon választott tárgyszó és különösen a formai tárgyszó gyakorisága is. Az OSZK nagyjából a középmezőny alsó

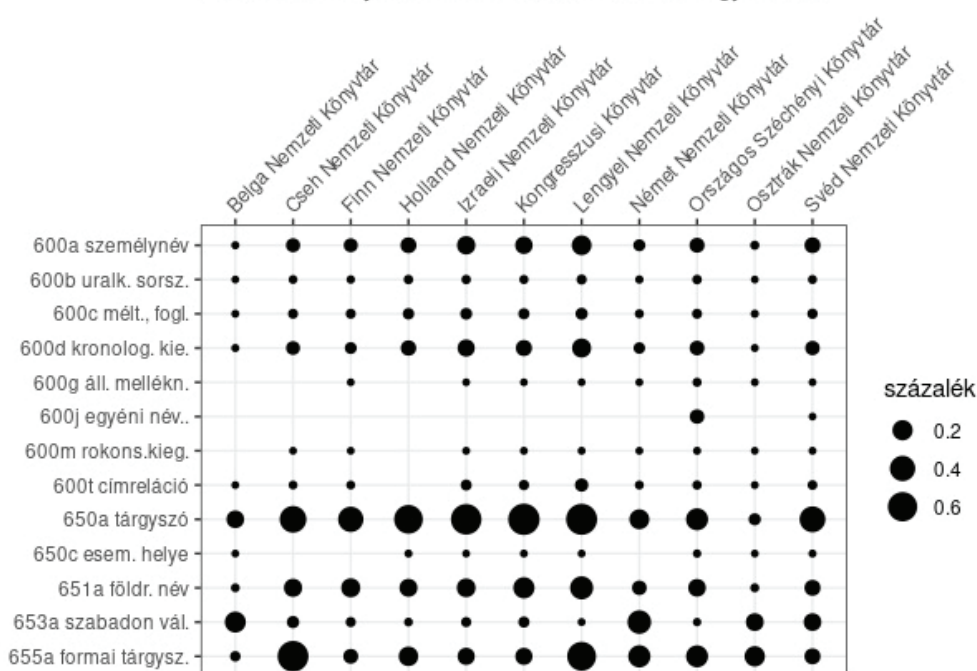
részében foglal helyet, almezők kitöltésének dolgában pedig feltűnően jó. Ez nem rossz helyezés, noha lennének egyszerűen megoldható javítandók, mint a \$2 forrásadatok, a \$0 és a \$1 azonosítók megadása, továbbá a formai tárgyszavak gyakoribb használata.

Összegzés: az OSZK helye

Az OSZK a tárgyalat 00X és 6XXX adatmezők feldolgozásának dolgában a középmezőnyben, annak az alsó részében foglal helyet. Ez – tekintve a Nemzeti Könyvtár nemzetközi viszonylatban gyenge

Szaktárgyszavak

A rekordok hány százalékában található szaktárgyszavak?



13. diagram A 6XX mezők fontosabb almezői összevont buborékdiagramban, a jobb áttekintés érdekében

anyagi lehetőségeinek fényében – inkább meglepő. Összefügghet azzal, hogy munkatársai révén egyelőre jelentős hozzáértési szellemi tőke halmozódott fel. A könyvtár ezt éli fel.

A 00X mezők kitöltését alapértelmezéseket tartalmazó „sablonok” alapján végzik. Ezt e tanulmány felismerései alapján ideje lenne fölülvizsgálni. Az alapértelmezéseket maga a könyvtári rendszer tartalmazza, nem rutin feladat a kibogarászásuk. Noha a többi nemzeti könyvtár alapértelmezései se közismertek (csak áttételesen, az adatbázisaik elemzése alapján állapíthatók meg), úttörő munka lenne az OSZK alapértelmezéseit az indoklásukkal nyilvánosságra hozni, angol fordításban is.

Nemzetközi összehasonlításban is kiemelkedő az aprónyomtatvány-, zenemű- és térképtári feldolgozó munka. Együtt a könyvek monografikus és az időszaki kiadványok feldolgozásával, mindegyikre kitüntetetten jellemző, hogy közös tárgyszórendszert, a Köztauruszt használják. Ez nemzetközileg az egyik legnagyobb névtér, tezaurusz a maga közel 150.000 lexikai egységével, melyek szemantikai szócikkekben vannak rendezve, és jelentős részük magyarázattal, forrás- és egyéb megjegyzésekkel rendelkezik. A Köztaurusz a századfordulón az összes online elérhető magyar tezaurusz és tárgyszójegyzék egyesítésével és egységesítésével készült és követő tartalmi karbantartása egyelőre biztosítva van [6]. Egyelőre – sajnos – csak professzionális használatra alkalmas felhasználói felülettel rendelkezik. Mint ilyen, bármely magyarországi könyvtár használhatja, amivel sok helyen élnek is. Szoftverének (a Relexnek) a sorsa nincs még megnyugtatóan rendezve [7].

Összegzés: két speciális új MARC21 metaadat-azonosító

A MARC21 a mesterséges intelligencia jövőbeli alkalmazhatóságát előkészítendő, bevezetett két új metaadat-azonosítót. Mindkettő feloldató (dereferenciális) azonosító, azaz megadják a metaadat forrását.

\$0 metaadat-azonosító (authority record control number or standard number)²³

- A \$0 almező tartalmazza a kapcsolódó besorolási adat(rekord) rendszerellenőrző számát, vagy egy szabványos azonosítót. Ezek az azonosítók lehetnek szövegesek vagy egységes

erőforrás-azonosítók (URI). Ha az azonosító szöveges, akkor az ellenőrzőszámot vagy azonosítót a besorolási rekordot kezelő MARC szervezeti kód, szabványosított metaadatséma esetén pedig a metaadatséma forráskódja előzi meg kerek zárójelben. Ha az azonosítót webes lekérési protokoll, például HTTP URI formájában adják meg, akkor nem kell a megelőző zárójel.

- A \$0 almező tartalmazhat egy URI-t, amely egy entitás nevét vagy címkéjét azonosítja. Ha feloldató (dereferenciális) az URI, akkor a nevet leíró információra mutat. Az az URI, amely nem az entitás nevét, hanem az entitást magát közvetlenül azonosítja, a \$1 valós fizikai tárgy almezőjébe kerül

A \$1 valós fizikai tárgy azonosítója (real world object)

- A \$1 almezőben szereplő URI terjedelmet, előfordulást azonosít, mint például dolgot, valós fizikai tárgyat. Lehet aktuálisan létező (például egy konkrét sisakveret) vagy fogalmi természetű (például egy adott államban meghatározott időben létező vagy létezett politikai rendszer).
- Ha az URI feloldató (dereferenciális), akkor a tárgy, dolog leírására mutat. A tárgy, dolog nevét (metaadatot) azonosító URI a \$0 almezőbe kerül.

A nemzeti könyvtárak jelentős része már elkezdte ezeknek az azonosítóknak a használatát. Legfőbb ideje, hogy erre az OSZK-ban is sor kerüljön.

Összegzés: a nemzeti könyvtárak általános gyakorlata

A rekordfej és a vezérlőmezők (00X) értelmezése, alkalmazása minden jel szerint további pontosítást, fejlesztést igényel. Jelenleg rendkívül széttartó a gyakorlat. Hogy pontosan milyen, az a hatalmas adatmennyiség miatt csak szűrőpróbaszerűen ellenőrizhető. Az egyes nemzeti könyvtárak szakemberei azonban, ismerve a saját gyakorlatukat, mindezek alapján jobban tudják értelmezni mindazt, amire ebben a tanulmányban talán sikerült jelzesszerűen utalni. És ezen a nyomon haladva lehetőség nyílhat a katalógusaik javítására, és a MARC21 fejlesztésére is.

Jó volna, ha nyilvános lenne, hogy mely könyvtárakban áll rendelkezésre a rekordfej és a vezérlő mezők (00X) kitöltéséhez dokumentált szabályzat, és hol dolgoznak pusztán szóbelileg közvetített gyakorlat alapján. Szakmai nyilvánosság hiányában ezek a gyakorlatok se nem egységesülhetnek, se nem termékenyíthetők meg a szükséges fejlesztéseket.

Az elemzés fontosabb konkrét tapasztalatai:

- A rekordfej és a vezérlő mezők (00X) pozícióin a dokumentumok típusainak egykarakteres mesterséges nyelvű osztályozási jelzeteit rögzítik. Ezek a karakterek pozícióként elvileg meghatározott szempont szerinti tipológiát képviselnek. Az egyetlenszer alkalmazható, egyetlen pozícióhoz kötött, egykarakteres osztályozási jelzetrendszer csak nagyon általános, durva besorolást tesz lehetővé.
- Az egyes osztályok meghatározásai elégtelennek tűnnek. Ez az összehasonlításokban derül ki. Például nem látszik teljesen világosan, hogy mi értendő alárendelt részegységnek (Rf/07=d) kéziratok vagy plakátok esetén (Rf/06=d, ill. =k). A kitöltési szabályzatok ismerete erre talán kitér, de hozzáférhetetlenek.
- A rekordfej pozícióinak kitöltése kötelező, a 00X vezérlő mezők kitöltése nem kötelező. Ennek következtében válhatott lehetővé, hogy az utóbbi mezők feldolgozottsága rendkívül kicsi. Annyira, hogy nem is lehet érdemben részletesen elemezni a feldolgozásukat, és erről érvényesebb észrevételeket tenni.
- A pozíciók kitöltését feltehetően helyi szabályzatok vagy szóbeli gyakorlat határozza meg. Ez a gyakorlat explicit szakmai nyilvánosságának hiányában nagyban korlátozza az elemzésben felvetődő problémák átfogó, konszenzusos megoldását, nehezíti a nemzetközi gyakorlat egységesülését. Hiányzik a helyi szabályzatok szakmai közérdeket szolgáló nyilvánossága és megtárgyalása, hiányzik a rekordfej és a vezérlő mezők feldolgozásáról szóló beszámoló. A MARC Tanácsadó Bizottság (MARC Advisory Committee) honlapján (<https://www.loc.gov/marc/mac/index.html>) mutató segítségével jól követhető például a tételfej és a 00X adatmezőkkel kapcsolatos változtatási előterjesztések és eredményeik. Ezek mind jelzetek (kódértékek)

tartalmának bővítésére, pontosítására, nevének megváltoztatására, néhány új érték fölvételére (hangzó anyagok, mozgókép és számítógép által kezelt típusok) vonatkoznak. Van néhány javaslat a 008 és a 047-es mezők jobb összehangolására. Átfogó átdolgozásra irányuló igénynek, a 00X tartalmi típusrendszerének kritikájára vonatkozó törekvéseknek (egyelőre) nincs nyoma.

- Mindez azt is jelenti, hogy nem állnak rendelkezésre hozzáférhető információk a 00X adatmezőkkel kapcsolatos MARC21 fejlesztésekhez sem.
- A felvetődő problémák és megoldásuk nincs ellentmondásban azzal, hogy lehetséges egy, a jelenleginél jobban kidolgozott, átfogó tipológiai osztályozás a vezérlő mezőkben (007 és 008), nem utolsósorban annak érdekében is, hogy a katalógusok keresőrendszereit az ezekben a mezőkben való, felhasználóbarát keresésre felkészítsék.
- Sok esetben a tételfej dokumentumtípusával, illetve a 007 mező dokumentumtípusaival párhuzamosan léteznek szöveges mezők is (mint a 254, 256, 300, 513, 516, 588), melyekben ugyancsak osztályozhatók dokumentumtípusok, és a 655-ös mezőben kötött tárgyszavakkal is. Nincs a MARC21-ben szabályozva ezeknek a mezőknek a viszonya, melyiket milyen értelemben célszerű használni a redundancia csökkentése, a feldolgozás egyértelműsége érdekében. Ráadásul: szemben a tételfej és a 00X mezők kötött osztályozási jelzeteivel, a felsorolt megjegyzésmezők szövegesek, és a katalógizáló rendszerek keresői nincsenek egyáltalán felkészítve arra, hogy felhasználóbarát módon bevonhatók legyenek laikusok által is a keresésbe.
- A vezérlőmezők pozícióin kis számban, de sok katalógusban fordulnak elő definiálatlan értékek, nagyon kis számú találattal. Ezek sokszor nem kellően ellenőrzött tömeges importból származhatnak, és az általunk alkalmazott QA catalogue segítségével könnyen felismerhetők és kiszűrhetők.
- A 00X adatmezők szerinti könnyű kereshetőséget be kell építeni a katalóguskezelő rendszerek keresőfelületeibe.

version	records	all issues		issues in documented data elements		change	
		without	with	without	with	in records	in %
2020-11-21	18,787,911	0.00%	100.00%	13.60%	86.40%		
		0	18,787,911	2,555,640	16,232,271		
2022-03-14	19,154,445	75.93%	24.07%	75.98%	24.02%	+11,997,544	+62.376%
		14,544,099	4,610,346	14,553,184	4,601,261	↑	↑

4. ábra Javítási grafikon a Brit Nemzeti Könyvtárban (BL)

- Végül az egyik legfontosabb: a katalógusok, repozitóriumok stb. metaadat-sémáiba be kell építeni az adott intézményben vagy felhasználói közösségben kialakult/kidolgozott egyedi használati szokásrendjének (alkalmazói profiljának, application profile) tárolására és kezelésére alkalmas részt. Mégpedig úgy, hogy azon keresztül a katalogizálási rendszerbe beépített alapértelmezéseket a könyvtáros felhasználó tudja módosítani.

Kitekintés

Hogy a QA catalogue által feltárt katalógusokban mennyire eredményes lehet a hibajavítás, azt például a BL katalógusállományában elvégzett munka igazolja.

Ha a rekordban akárcsak egyetlen hiba van, akkor az hibás rekordnak számít. Két eshetőséget különböztetünk meg:

(1) az adat értéke egyértelműen eltér egy jól dokumentált előírástól. Például, ha a 245\$ind1 értéke térköz, akkor az nem felel meg a MARC21 előírásának, mivel ennek az indikátornak 0-t vagy 1-et kell tartalmaznia.

(2) lehet továbbá olyan adatelem, amely a MARC21-ben vagy annak ismert kiterjesztésében nincs definiálva, ezért nem ítéhető meg az érvényessége, mert nem ismerjük a szabályait.

Például: 024\$9 esetében nem vagyunk biztosak benne, hogy probléma-e, lehet, hogy ez egy lokális almező, mert a 9-es erre van fenntartva (de lehet, hogy elírás is). 250\$p esetében szintén nem vagyunk biztosak; a MARC21-ben nem szerepel, de mivel nem 9, sokkal kisebb a valószínűsége annak, hogy ez egy lokális almező. Mindkét esetben előfordulhat, hogy a rekordot máshonnan importálják, és a lokális almező meghatározása az eredeti intézményben elérhető, de az aktuális katalógusban már kilóg a kontextusból.

A baloldali oszlop fekvődiagramjai az összes, tehát az (1) és a (2) eseteket mutatják meg. Ezek szerint az első vizsgálat idején (ez a felső sor) a rekordok 100%-ában volt adatmezőt érintő hiba. Ezek 70%-ban indikátor-hiba, közel 50%-ban pedig vezérlő mező hiba. További hibatípusok a vezérlő mezők (00X) esetében az érvénytelen adatértékek, az XXX adatmezők esetében a definiálatlan mezők, az indikátorok esetében a nem üres (kódolatlan=) indikátorhelyek. Ezen felül a rekordok 17%-ban a definiált mezőkben volt definiálatlan almező. Az alsó sorban a formális hibajavítások javításának az eredménye látható fekvődiagramban.

A második oszlop fekvődiagramjai csak az (1) típusú hibaesetekre vonatkoznak.

Irodalom

- [1] HUNMARC *A bibliográfiai rekordok adatcsere formátuma. KSZ 4/1:2002*, Könyvtári és Szakirodalmi Tájékoztatási Szabványosítási Bizottság, Országos Széchényi Könyvtár, Budapest, 2002.
- [2] Király, P., Snyk, B., Takács, Á., Svetlik, R., Virolainen, T., Heggø, D. M. O., Rolschewski, J., Kelly, M., Hemme, F., Voß, J. *QA catalogue: Release v0.6.0.*, 2022.
<https://doi.org/10.5281/zenodo.7347755>
- [3] Király, P. *Validating 126 million MARC records*, DATeCH2019 Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage, Brussels, Belgium, 2019. p. 161–168.
<https://doi.org/10.1145/3322905.3322929>
- [4] Király, P. *Empirical evaluation of library catalogues*, EuropeanTech Newsletter, 15, 2020. Elérhető: <https://pro.europeana.eu/page/issue-15-swib-2019#empirical-evaluation-of-library-catalogues>
- [5] Network Development and MARC Standards Office *MARC21 Format for Bibliographic data*, Library of Congress, Update No. 30, 2020. Elérhető: <https://www.loc.gov/marc/bibliographic/>
- [6] Ungváry, R. *Az OSZK tezaurusz és a KÖZTAURUSZ*, Könyvtári Figyelő, 47(1), p. 11–40, 2001. Elérhető: <https://epa.oszk.hu/00100/00143/00037/index.htm>
- [7] Styles, R., Ayers, D., Shabir, N. *Semantic MARC, MARC21 and the Semantic Web*, Proceedings of the WWW2008 Workshop on Linked Data on the Web. CEUR, 2008. Elérhető: <https://ceur-ws.org/Vol-369/paper02.pdf>
- [7] Ungváry, R. *Tezauruszkezelő programok és a RELEX*, Tudományos és Műszaki Tájékoztatás, 48(1), p. 3–15, 2018. Elérhető: <https://tmt.omikk.bme.hu/tmt/article/view/2097>
- [8] Ungváry, R. *Besorolási, szabványosított, normatív vagy „autorizált”*, Tudományos és Műszaki Tájékoztatás, 66(6), p. 328–342, 2019. Elérhető: <https://tmt.omikk.bme.hu/tmt/article/view/12309>
- [9] Ungváry, R. *MARC21 tartalmi adatmezők használata jelentősebb nagykönyvtárakban. Egy elemzés néhány tanulsága*, Networks-hop 2020, Pécs, 2020, p. 33–53.
<https://doi.org/10.31915/NWS.2020.4>
- [10] Ungváry, R. *Ismeretszervező-könyvtári rendszerek tartalmi feltárásának összehasonlító vizsgálata MARC21 környezetben*, Tudományos és Műszaki Tájékoztatás, 67(11), p. 655–680, 2020. Elérhető: <https://tmt.omikk.bme.hu/tmt/article/view/12776>
- [11] Ungváry, R., Király, P. *Bemerkungen zu der Qualitätsbewertung von MARC-21-Datensätzen*, In: Qualität in der Inhaltserschließung, Bibliotheks- und Informationspraxis, Volume 70, Edited by: Franke-Maier, M., Kasprzik, A., Ledl, A., Schürmann, H. De Gruyter, 2021, p. 177–227.
<https://doi.org/10.1515/9783110691597-011>
- [12] Ungváry, R., Király, P. *A MARC21 tételfejének és kódolt tartalmi jellemzőinek feldolgozási minősége néhány nemzeti könyvtárban. Egy elemzés tanulságai*, Tudományos és Műszaki Tájékoztatás, 69(7), p. 293–313, 2022.
<https://doi.org/10.3311/tmt.13174>

Végjegyzet

- 1 Az OSZK-ban a HUNMARC-ot használják, de az összehasonlításra bevont mezők esetében ezek azonosak a MARC21 mezőivel.
- 2 Letölthető: <http://libris.kb.se/xsearch>
- 3 Az adatokat a könyvtártól csak kutatási céllal kaptuk meg
- 4 Letölthető OAI-PMH-n keresztül a https://eu02.alma.exlibrisgroup.com/view/oai/43ACC_ONB/request helyről
- 5 Letölthető: <http://data.nationallibrary.fi/download/>
- 6 Letölthető SPARQL (<https://libris.kb.se/sparql>) és az Xsearch API-k (<http://libris.kb.se/xsearch>) kombinálásával
- 7 Letölthető: <https://data.bn.org.pl/databases>
- 8 Letölthető OAI-PMH és Z39.50 API-k kombinálásával. Az API-k végpontjait csak kutatási céllal használhattuk.
- 9 Az adatokat a könyvtártól csak kutatási céllal kaptuk meg.
- 10 Az adatokat a könyvtártól csak kutatási célra kaptuk meg.
- 11 Az adatokat az OAI-PMH szerver (<http://services.kb.nl/mdo/oai>) segítségével arattuk le.
- 12 A könyvtár rekordjai letölthetőek a <https://www.en.nkp.cz/about-us/professional-activities/open-data> helyről.
- 13 Az adatok a <http://ddb.qa-catalogue.eu/oszk/> helyről nyilvánosan elérhetők.
- 14 A MARC21 2018-as angol változatról készült egy „MARC 21: a bibliográfiai rekordok adatcsere-formátuma–MUNKAANYAG” című fordítás, mely csak az adatmezők, az almezők és indikátorok fordítását tartalmazza (ford. Kasza Zsófia.) <http://www.oszk.hu/sites/default/files/MARC21_bibliografiai_rekordok_adatcsere_formatuma.pdf>, és nem KSZ kiadás. A kéziratban nincs jelezve, hogy lektorálták volna, a fordításhoz egyáltalán nem vették figyelembe a már meglévő, 2002-es KSZ/4.1 fordítási megoldásait, a kézirat nyilvános vitájából kizárták azokat, akik mindezt megpróbálták szóvá tenni, a mező- és almezőnevek fordítása mereven követi az angol kifejezéseket. A benne alkalmazott magyar terminológiai problémákról lásd [15].

- 15 „Collection of digitized photographs of buildings contributed by SBSE members, formerly available as a subset of slides under the SBSE Slide Exchange Program. Slides date from October 1985 to the present, and are intended to be used as a teaching resource by building science educators. Includes image indexes and author annotations accompanying the slides.”
- 16 Többször megkíséreltünk levélben erre az Osztrák Nemzeti Könyvtárban (ÖNB) rákérdezni, de érdemben nem válaszoltak
- 17 Ez nem a szűkebb értelemben vett katalógus, hanem a teljes könyvtáré: <https://www.loc.gov/search/>
- 18 <https://www.loc.gov/search/?in=&q=Esterh%C3%A1zy&new=true> Utolsó elérés: 2023. 06. 20. Azóta a találatok száma bővült.
- 19 OSZK: Rokonsági kiegészítő (id,, ifj., jelző, ill. ezek nyelvi változatai.
- 20 Styles et al. [7]
- 21 Az „authority data” magyar megfelelőjeként a „besorolási adat” kifejezést használjuk, és nem az angol név szolgái átvételét, az „autorizált adat” nevet [8].
- 22 A tezaurusz adatai egyben a KÖZTAURUSZ részei is.
- 23 A meghatározások a hivatalos MARC21 kiadvány angol szövegét követik [5].

Beérkezett: 2023. június 26.



Ungváry Rudolf
informatikus mérnök
E-mail: ungvary@gmail.com



Király Péter
szoftverfejlesztő, kutató
Göttingen eResearch Alliance
E-mail: peter.kiraly@gwdg.de